HIO STATE
NIVERSITY
AUG 15 1961
LIBR
EDUCATION LIBRARY

VOL. 3, NO. 2, JULY, 1961

human factors

THE JOURNAL OF THE HUMAN FACTORS SOCIETY

Contents

THE EFFECT OF THE PRECISION OF INFORMATIONAL FEEDBACK ON HUMAN TRACKING PERFORMANCE 77
Darwin P. Hunt

DEVELOPMENT OF DESIGN CRITERIA FOR INTELLIGENCE DISPLAY FORMATS 86
W. D. Hitt, H. G. Schutz, C. A. Christner, H. W. Ray and J. L. Colley

A COMPARISON OF VERTICAL AND HORIZONTAL ARRANGEMENTS OF ALPHA-NUMERIC MATERIAL (EXPERIMENT I) 93
John L. Colley

AN EVALUATION OF FORMATS FOR GRAPHIC TREND DISPLAYS (EXPERIMENT II) 99
Howard G. Schutz

AN EVALUATION OF METHODS FOR PRESENTATION OF GRAPHIC MULTIPLE TRENDS (EXPERIMENT III) 108
Howard G. Schutz

AN EVALUATION OF FIVE DIFFERENT ABSTRACT CODING METHODS (EXPERIMENT IV) 120
William D. Hitt

AN EVALUATION OF THE EFFECT OF SELECTED COMBINATIONS OF TARGET AND BACKGROUND CODING OF MAP-READING PERFORMANCE (EXPERIMENT V) 131
Charlotte A. Christner and Horace W. Ray

THE AUTHORS 147

PERGAMON PRESS NEW YORK • LONDON • PARIS • LOS ANGELES

OFFICERS AND EXECUTIVE COUNCIL

The Human Factors Society is a non-profit California Corporation. The officers and members of the Executive Council are:

PRESIDENT: Stanley N. Roscoe, 1961
PRESIDENT-ELECT: Jack W. Dunlap, 1961
VICE-PRESIDENT: Wesley E. Woodson, 1961
PAST-PRESIDENT: Arnold M. Small, 1961
SECRETARY: Paul G. Cheatham, 1961
TREASURER: Donald W. Conover, 1961
EDITOR-IN-CHIEF: Stanley Lippert, 1959–1961

EXECUTIVE COUNCIL: Charles I. Barron, 1960–1962; Max W. Lund, 1959–1961; John Lyman, 1960–1962; Ross A. McFarland, 1959–1961; John J. Swearingen, 1961–1963; Lawrence Slote, 1961–1963

EDITORIAL BOARD: Walter P. Grether; Blaine H. Levedahl; J. C. R. Licklider; Stanley Lippert; John Lyman; John P. Meehan; Allison K. Simons; Ezra V. Saul; Neil D. Warren; Joseph W. Wulfeck

HISTORIAN: Donald T. Hanifan

Address general correspondence to Human Factors Society, 4113 Lee Highway, Arlington 7, Virginia, Attention: **Paul G. Cheatham**, Secretary, for Society business including payment of dues.

Address matters pertaining to the Journal or Bulletin to **Human Factors**, Biotechnology Laboratory, Department of Engineering, University of California, Los Angeles 24, California, Attention:

John Lyman, Managing Editor, for Journal or Bulletin Subscriptions or other publication business, including changes of addresses.

Stanley Lippert, Editor-in-Chief, for submission of Journal articles and letters to the editor.

Hilde Groth, Bulletin Editor, for Bulletin news.

COPYRIGHT © 1961
by the HUMAN FACTORS SOCIETY

PUBLISHED BY
PERGAMON PRESS, 122 E. 55th St., New York 22, N.Y.
Headington Hill Hall, Oxford, England.

Printed in Great Britain.

SUBSCRIPTION PRICE: MEMBERSHIP in Society includes subscription to both publications.

Journal: **NON-MEMBERS SUBSCRIPTIONS TO BE SENT DIRECT TO THE PUBLISHER, PERGAMON PRESS.** Commencing with Volume 2 the subscription price will be:
$10.00 (£3 10s.) per annum, for individuals who certify that the journal is for their personal use. $20.00 (£7) per annum, for libraries, government establishments and research institutions.

Bulletin: $5.00 per year. Orders to be sent to John Lyman.

The Effect of the Precision of Informational Feedback on Human Tracking Performance[1]

DARWIN P. HUNT, Behavioral Sciences Laboratory, Wright Air Development Division.

Research was conducted to determine human operator abilities in a compensatory tracking task as a function of the precision of the displayed error information for various levels of task difficulty and practice. The controller dynamics were not varied and were such that displacements of the control were integrated twice before affecting the tracking error. Tracking error and the extent of control motion were the criterion measures. The results suggest that (a) increases in the precision of displayed error information result in a negatively accelerated reduction in the tracking error; (b) as the operator becomes more proficient, the relative superiority of an increased precision of displayed error is reduced; (c) the effects of the precision of information on control motion depends upon the difficulty of the task; and (d) the amount of control motion generally decreases as the operator becomes more proficient and the magnitude of this decrease is greater for an "easy" task than for a "difficult" task.

INTRODUCTION

Man, as a component in a man-machine system, can perform his tasks only as effectively as the information made available to him will permit. In some cases, a simple indication of whether a system parameter is within certain limits, i.e. a "go—no-go" display, is sufficient. Other situations require the display of information concerning not only whether the parameter has exceeded the limits, but also the direction of this excess. In still other tasks, information related to the existence, direction, and magnitude of an out-of-limits condition must be displayed to the operator if an acceptable level of over-all system performance is to be attained. In this latter situation, a question that arises is, "With what degree of precision (resolution or specificity) should the information concerning 'error' magnitude be presented so that optimal performance will be realized?"

At the present time, the degree of specificity of the information presented to the man is determined largely by the availability and reliability of various display techniques; the weight, size, and economic penalties involved, and the rational determination of the utility of the information to the operator. If the abilities of man are to be used most effectively in future systems, a more sound basis for the consideration of human factors in making this design decision must be provided. The purpose of this paper is to contribute knowledge to assist the designer of man-machine systems in determining the degree of precision of feedback information to display to the man in order to obtain the optimum output of the total system. More specifically, this research was designed to determine the effects of feedback precision on tracking performance and how the effects of precision are influenced by practice and the difficulty of the tracking task.

METHOD

Subjects

Sixty-four male undergraduate college students were paid to serve as subjects. None had any apparent visual or motor deficiencies.

Apparatus

The apparatus, which is represented in Fig. 1, consisted essentially of a target course

[1] This paper is based on research conducted at the Behavioral Sciences Laboratory, Aerospace Medical Division, Wright-Patterson Air Force Base, Ohio, and is reported in more detail in WADC Technical Report 58-584. The comments of Dr. John P. Hornseth are gratefully acknowledged.

generator, an error transformer, a Cathode Ray Tube display, a lever control, and an analogue computer.

The target course generator was designed so that either of two target courses could be selected for presentation on the CRT. The two target courses differed only in terms of frequency of motion. The "easy" target course was constructed by combining $\frac{2}{3}$, $1\frac{1}{3}$, and $3\frac{1}{3}$ cpm sine waves, the $3\frac{1}{3}$ cpm component being 180° out of phase. This target course was traversed once during each trial of 90 sec duration. To obtain the "difficult" target course, the "easy" course was traversed three times per 90 sec. Thus, the "difficult" course consisted of a combination of 2, 4, and 10 cpm sine waves.

The instantaneous difference between the output of the target course generator and the output of the control system, i.e. the tracking error, was fed into the error transformer. This tracking error (true error) was then transformed in accordance with the particular condition of feedback specificity under which a given subject was scheduled to perform (see Fig. 2a, b, c, and d). The output of the error transformer was then fed directly to the CRT display.

The subjects display consisted of a 5-in. Cathode Ray Tube with a P–11 (low persistance) phosphor. The front of the CRT, except for a slot 1 in. high by $4\frac{1}{2}$ in. wide, was covered by a piece of black cardboard. The transformed tracking error was displayed within the slot along the horizontal dimension. The zero error position was indicated by two vertical lines $\frac{1}{4}$ in. apart; these lines were centered horizontally in the slot

The control used by the subject was a lever 26 in. in length, with a rubber grip on the top portion. The lever could be moved to the left and right but was fixed in the fore-aft direction.

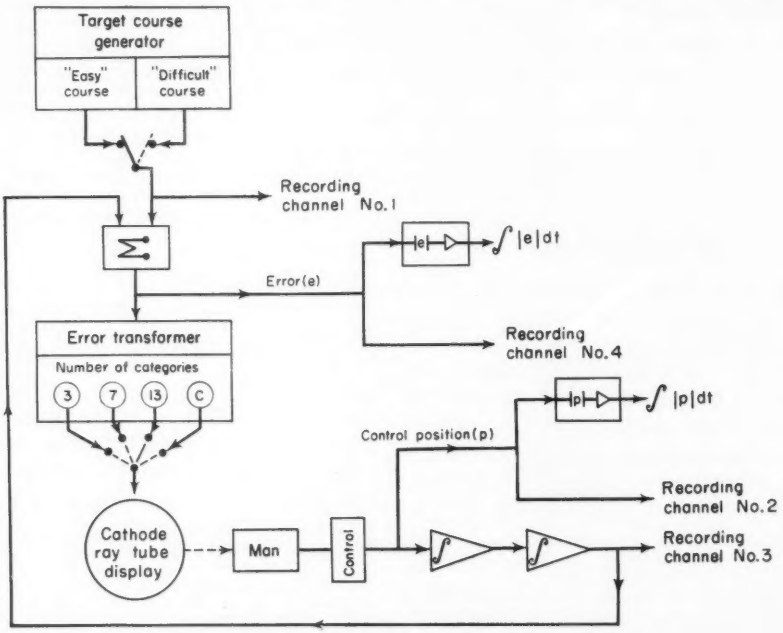


Fig. 1. Block diagram of the experimental tracking apparatus.

When no force was applied in a sideward direction, spring centering returned the lever to the center upright position at which the output of the control was effectively zero. The base of the control was fixed so that when the control was in the neutral position, the top of the lever was directly in front, 15 in. below and 15 in. away from the center of the display.

The computer was used to insert controller dynamics into the system and also to obtain criterion measures. Attached to the subject's control lever was a potentiometer. Across the potentiometer was a voltage, whose sign and magnitude depended upon the direction and extent of displacement of the control lever from the center upright position. This voltage was fed through two integrating amplifiers, in series, of the computer. The integration rates were such that

$$(OUT) = 1.0885(IN)t^2$$

where OUT represents the output of the second integrating amplifier, IN is the voltage across the potentiometer of the control lever, and t is time.

Voltages proportional to tracking error (e) and control position (p) were obtained at the points indicated in Fig. 1. These voltages were fed into the computer which was programmed so that voltage analogues of the integrated absolute value of the tracking error ($\int|e|\,dt$) and the integrated absolute value of the control position ($\int|p|\,dt$) could be obtained for each trial.

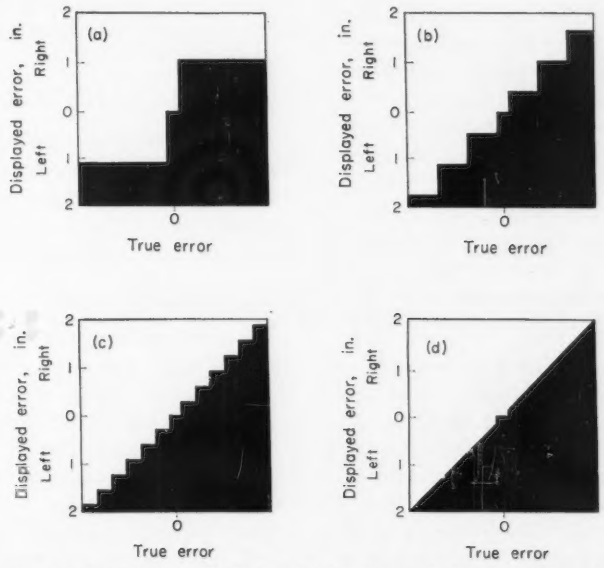


Fig. 2. The four error transformations studied. The displayed error is shown in inches of displacement of the dot on the subject's display.

In order to maintain a relatively constant subject environment, the display, control lever, and chair were enclosed in a 6 ft × 6 ft × 6½ ft sound shielded booth. Illumination was provided by a 3-W neon light located at the top of the wall behind the subject. A fan mounted in an opening in the ceiling of the booth provided adequate air circulation, and the noise made by the fan masked out cues such as the clicking of scoring counters.

Procedure

Two levels of task difficulty were studied. These corresponded to the "easy" and "difficult" frequencies of target course motion described above. Four degrees of specificity of feedback information were investigated (see Fig. 2).

(a) Three categories—only information concerning the existence and direction of the error was displayed to the subject (see Fig. 2a).

(b) Seven categories—information concerning the direction of the error and three indices of error magnitude in either direction were presented to the subject (see Fig. 2b).

(c) Thirteen categories—similar to (b) except that six indices of the magnitude of the tracking error in either direction were presented to the subject (see Fig. 2c).

(d) Continuous—the error dot moved "smoothly" across the CRT display in either direction from the zero error position in direct relation to the tracking error (see Fig. 2d).

The sixty-four subjects were randomly assigned, eight to each of the eight combinations of task difficulty and feedback specificity. The subject performed under his assigned condition for two sessions which were held on successive days, six subjects being tested each day. In order to minimize any biasing effect of random variability resulting from factors such as uncontrolled environmental conditions and equipment variations, the following two precautions were taken: (a) none of the eight experimental conditions was represented more than once on any one day, and (b) each of the four degrees of feedback specificity was represented at least once on each day.

A session consisted of fifteen 90-sec trials, with a 45-sec rest period between each trial with one exception, *viz.* a 4-min rest period was given between trials 8 and 9. Prior to the first session, the subject was instructed with respect to the operation of the control, the controller dynamics, and the nature of the task; he was also given information concerning the trial and rest period durations and the activities permitted during the rest periods. These instructions were repeated to the subject immediately prior to beginning the second session.

A warning buzzer was sounded 5 sec before each trial. One second before the trial, the controller system was energized. At the end of the trial the buzzer was sounded again, and the subject then began his 45-sec rest period, during which he remained seated and was not permitted to smoke. During the 4-min rest, the subject relaxed outside the experimental booth.

RESULTS AND DISCUSSION

The two basic measures of performance were the integrated absolute error (Tracking Error Score) and the integrated absolute control position (Control Motion Score). These scores were obtained for each subject for each trial.

Tracking Error

The mean tracking error performance for each trial under each of the eight experimental conditions (four degrees of specificity of feedback information by two levels of task difficulty) was computed. These data are presented for the "easy" task in Fig. 3, and for the "difficult" task in Fig. 4.

First, inspection of these data suggest that an increase in the number of categories presented to the subject seems to result in an improvement in performance. Second, the relative improvement associated with an increased number of categories appears to differ depending upon the stage of practice. These and other relationships were tested by a modified Lindquist[4] Type III analysis.

The statistically significant effect of the number of categories ($p < 0.001$) supports the above-mentioned observation that increases in the number of categories result in improved performance. Inspection of Fig. 5 reveals that the tracking error decreases monotonically with increases in the number of informational categories and that the major reduction in tracking error is associated with increasing the

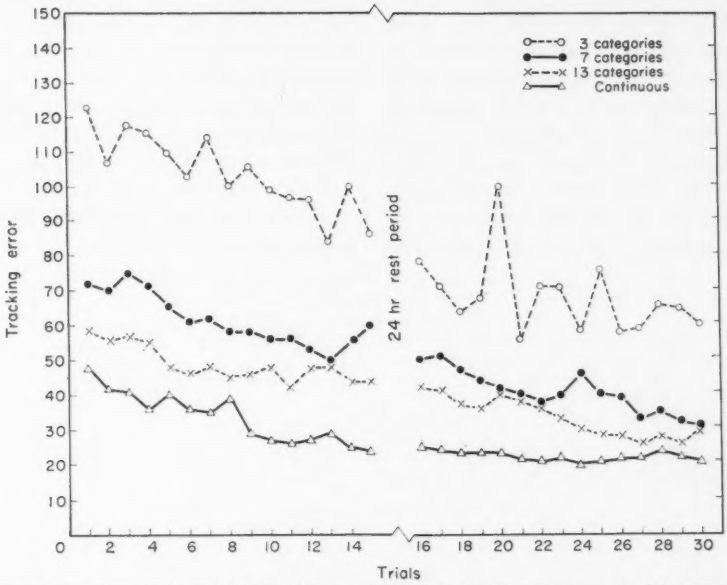


Fig. 3. Tracking error on the "easy" task as a function of practice. Performance on each of the four informational conditions is shown separately. Each point is the mean tracking error score for the eight subjects that performed under that condition.

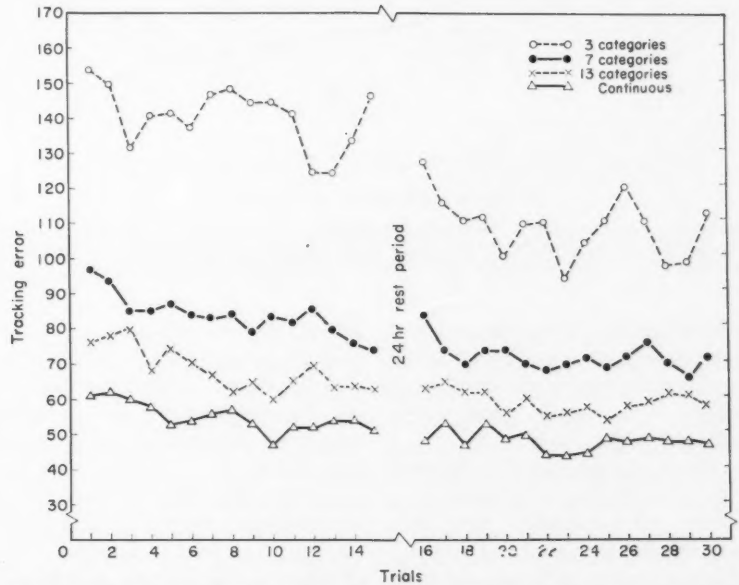


Fig. 4. Tracking error on the "difficult" task as a function of practice. Performance on each of the four informational conditions is shown separately. Each point is the mean tracking error score for the eight subjects that performed under that condition.

number of categories from three to seven. This improvement in performance realized by an increased number of informational categories was not influenced by the difficulty of the task as evidenced by the fact that the interaction of these factors is not significant. However, the tracking error associated with the "easy" task generally was less than the error associated with the "difficult" task ($p < 0.001$). The continuous condition. Although the design of the experiment does not permit a conclusive evaluation of tnis interaction, such a result could be accounted for in terms of the differential development of inhibitory effects for the different informational conditions.

It was originally expected, on the basis of work by Ammons,[1] Bilodeau,[2] Brown,[3] and Noble and Broussard[5] that the greater the

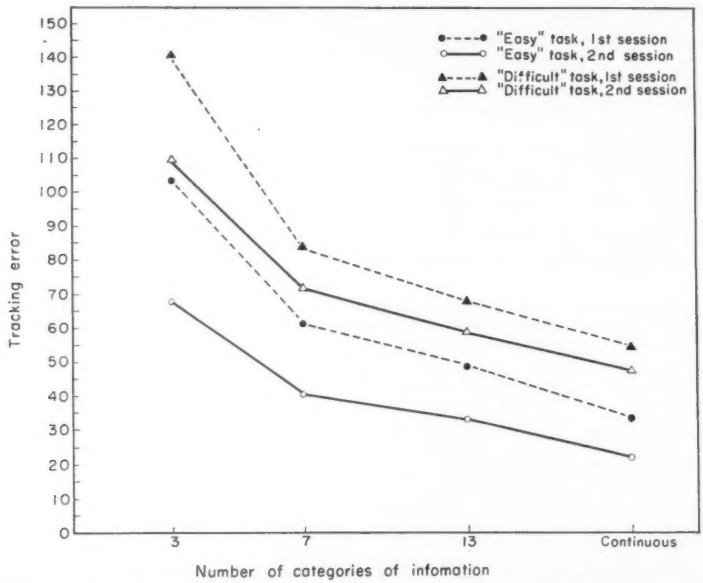


Fig. 5. Tracking error as a function of the number of categories of error information. Each point is the mean tracking error score for each session of the eight subjects that performed under each of the eight experimental conditions (2 levels of task difficulty × 4 degrees of feedback specificity).

lack of statistical significance between categories and trials, and among sessions, categories, and trials suggests that the rates of improvement—when averaged over sessions (and within sessions)—did not differ.

The main effects of sessions and of trials and the interaction between sessions and categories were significant (all $p < 0.001$). The latter interaction is reflected in Figs. 3 and 4 by the fact that the performance difference between Session I and Session II is relatively greater for the 3-category condition than it is for the number of categories, the more rapid the rate of improvement with practice. However, the findings of the present study fail to support this hypothesis; none of the informational conditions exhibited a rate of improvement any slower than that of the continuous condition. One could account for this result by introducing the notion of a "performance ceiling", i.e. the continuous informational condition permitted such a high initial level of performance that there was little room for improvement.

It can be concluded, rather confidently, that there was a reduction in the tracking error as the number of categories of error information was increased. Tracking error is a valid and widely used measure of performance for many practical situations.

Control Motion (CM)

It will be recalled that the subject's control system was such that displacements of the control lever were integrated twice before affecting tracking error, i.e. an acceleration control was used. One of the main purposes for this was to provide a task that would simulate the general type of controller situation with which man may be confronted in space travel in attempting to track manually some desired course in a single dimension. It is assumed that (a) there are only two outlets by means of which opposing thrusts, aligned with the dimension being controlled, may be delivered for guidance purposes, (b) the magnitude of the thrust is proportional to the extent of the displacement of the control lever, and (c) the vehicle being guided will tend to move in the direction of the displacement of the control lever from the center position. In such a situation, the effect of displacements of the control lever on the deviation from the desired course is approximated by a double integration of the lever displacement.

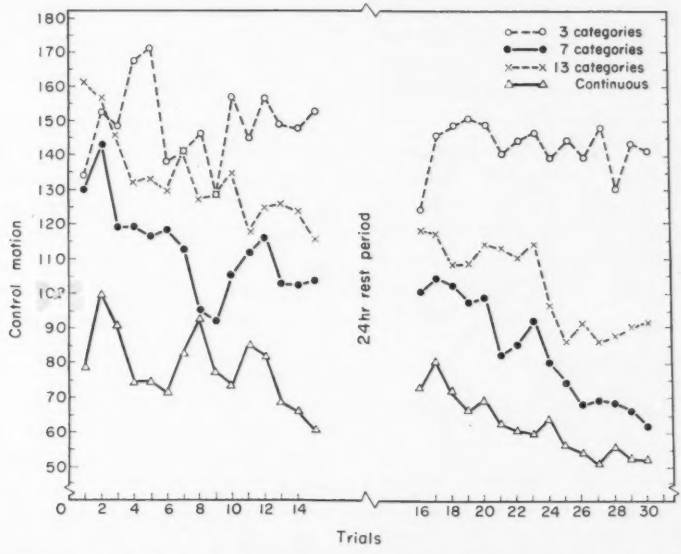


Fig. 6. Control motion on the "easy" task as a function of practice. Performance on each of the four informational conditions is shown separately. Each point is the mean control motion score for the eight subjects that performed under that condition.

Each time the control lever is displaced from the center position, a thrust is delivered which requires the expenditure of energy which is assumed to be stored in the vehicle. A factor that is presumably of interest to the designer of such reaction-controlled systems is the amount of fuel (or energy) consumed in the performance of the controlling task. In view of the reduced tracking error which is associated with increases in the number of error categories, it would be of considerable interest to determine the relationship between the

guidance energy expended and the number of error categories. It should be emphasized that the expenditure of energy of concern is that which is made available by the equipment, and not by the human operator. If the amount of energy expended is directly related to the number of error categories, then a trade-off between tracking error and guidance energy could be made. The CM Score provides an index of this rate of energy expenditure. The least control motion. This interaction between task difficulty and informational condition is statistically significant ($p < 0.05$). However, it should be noted that the CM Scores are not monotonically related to the precision of information, nor, indeed was the writer able to develop any rationale that would adequately account for the obtained relationships among the CM Scores, the informational conditions, and task difficulty. There is some evidence to

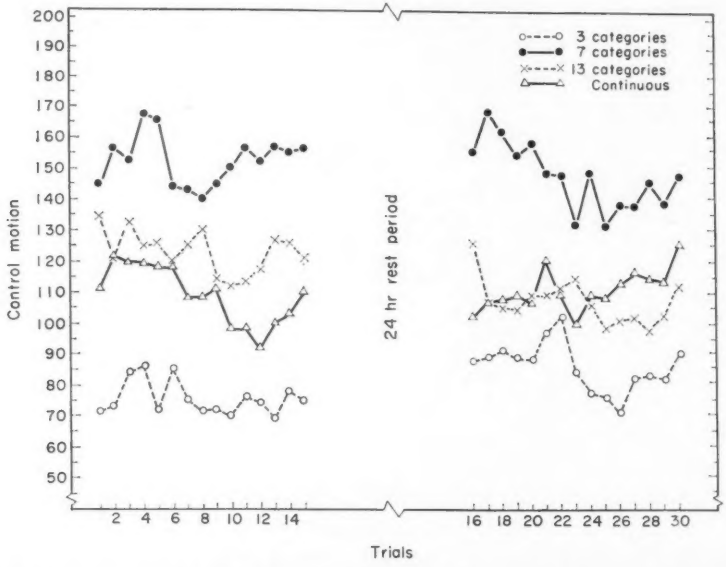


Fig. 7. Control motion on the "difficult" task as a function of practice. Performance on each of the four informational conditions is shown separately. Each point is the mean control motion score for the eight subjects that performed under that condition.

mean of the CM Scores for each trial under each of the eight experimental conditions was computed. These average values are presented in Fig. 6 for the "easy" task and in Fig. 7 for the "difficult" task.

Inspection of the curves of average control motion reveal that, for the "easy" task (Fig. 6), the continuous informational condition resulted in CM Scores that were consistently less than the others, while for the "difficult" task (Fig. 7), the three-category condition resulted in the indicate that the low amount of control motion under the three-category condition for the "difficult" task reflects a feeling of "futility" on the part of subject, i.e. the task was so difficult that the subject effectively "gave up".

Generally, the CM Scores seem to decrease as the subject is given more practice. The statistical analysis resulted in significance for both trials ($p < 0.001$) and sessions ($p < 0.01$), supporting this observation. Further, the CM Scores decreased more as a result of practice on

the "easy" task than they did on the "difficult" task (difficulty-by-trials, $p < 0.05$; difficulty-by-sessions, $p < 0.05$).

It also appears that there was a differential reduction in the CM Scores with practice depending upon the informational condition (Fig. 6), with the greatest rate of reduction being associated with the continuous condition. However, this notion was not supported by the statistical analysis; neither the information-by-trials interaction nor the information-by-sessions interaction was significant.

CONCLUSIONS

The purpose of this investigation was to provide knowledge concerning the relationship between human performance and the precision of feedback information.

Four degrees of feedback specificity and two levels of task difficulty were studied in terms of their effect on error and on control motion in the performance of a one-dimensional compensatory tracking task using an acceleration control. A given subject performed with 3, 7, or 13 categories of error information or with continuous error information. The "easy" task consisted of a combination of $\frac{2}{3}$, $1\frac{1}{3}$, and $3\frac{1}{3}$ cpm sine waves and the "difficult" task was composed of 2, 4, and 10 cpm sine waves.

Eight subjects were assigned randomly to each of the eight combinations of feedback specificity and task difficulty. Each subject performed thirty 90-sec trials under his assigned condition with 45-sec rest periods given between all trials except that a 4-min rest was given midway in each of the two sessions; 24 hr elapsed between sessions.

For tracking error, (a) performance improved monotonically, but by decreasing amounts, as the number of categories of information was increased, (b) this improvement was relatively unaffected by the difficulty of the task, and (c) the inhibitory effects on the first session were greater for the less specific informational conditions, i.e. the 3-category condition, than it was for the most specific informational condition, i.e. the continuous condition. The failure to find an optimal degree of specificity of informational feedback was attributed largely to the fact that the task was relatively complex. In terms of a control motion score, which was an approximation of the energy expended by the total system to perform the controlling task, it was found that (a) the effects of specificity of feedback information depended upon the difficulty of the task, (b) the scores generally decreased with practice, and (c) as practice proceeded, the scores decreased relatively more on the "easy" task than they did on the "difficult" task.

REFERENCES

1. Ammons, R. B., *Knowledge of Performance.* WADC Technical Report 54–14, Feb. 1954.

2. Bilodeau, E. A., Speed of Acquiring a Simple Motor Response as Function of the Systematic Transformation of Knowledge of Results. *Amer. J. Psychol.* **66,** 409–420, 1953.

3. Brown, J. S., *A Proposed Program of Research on Psychological Feedback (Knowledge of Results) in the Performance of Psychomotor Tasks,* pp. 81–87 in Research Planning Conference of Perceptual and Motor Skill, HRRC, Lackland Air Force Base, Conference Report 49–2, Nov. 1949.

4. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education,* New York, Houghton Mifflin, 1953.

5. Noble, C. E., and Broussard, I. G., Effects of Complex Transformations of Feedback upon Simple Instrumental Behavior. *J. exp. Psychol.* **50,** No. 6, 381–386, 1955.

Development of Design Criteria for Intelligence Display Formats[1]

W. D. HITT, H. G. SCHUTZ, C. A. CHRISTNER, H. W. RAY and L. J. COFFEY, Battelle Memorial Institute, Columbus, Ohio.

A broad research program to develop design criteria for intelligence display formats is outlined and the general findings of the program are discussed. Five specific areas of investigation are selected for study: a comparison of vertical and horizontal arrangements of alpha-numeric material, an evaluation of formats for trend displays, an evaluation of methods for presentation of graphic multiple trends, an evaluation of five abstract coding methods, and an evaluation of the effect of selected combinations of target and background coding on map reading performance. The five succeeding articles treat the five areas of investigation in detail.

INTRODUCTION

In the design of visual displays to be used in an intelligence center, the design engineer needs answers to a number of questions concerning human-factors problems. For instance, what types of formats should be used for the various displays? What types of coding methods should be employed? Is the use of color necessary? What is the best method for displaying trends, or historical data? The purpose of this research program was to attempt to provide answers to questions such as these. The present paper provides introductory and summary information describing the research program. The five following papers describe details of the various experiments.

The function of the intelligence system under study is to collect, process, display, and report pertinent military information about the enemy (or potential enemy). The aspect of this system with which the Battelle research program was primarily concerned is utilization of displays by "Center Control" personnel who exercise management over the data-processing activities. Broadly considered, Center Control exercises a filtering function with respect to directing the processing of system inputs and evaluating the quality of system outputs. An important requirement for visual displays is anticipated in connection with this function. Fig. 1 is a general representation of the system under study.

[1] This work was sponsored by the U.S. Air Force under Contract AF 30(602)-2078 (ref. RADC-TR-60-201).

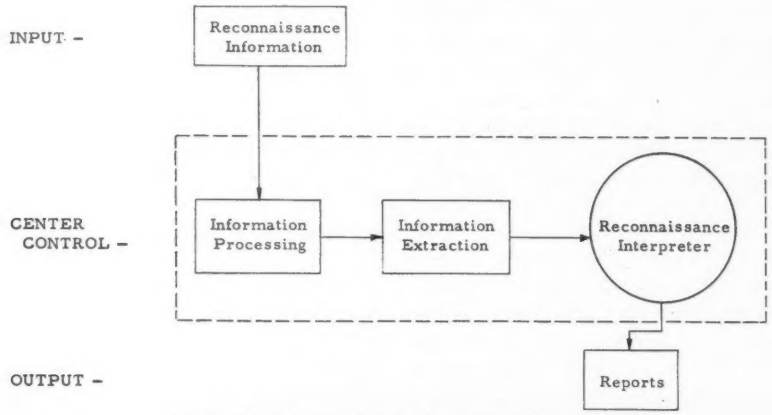


Fig. 1. General description of system under study.

DISCUSSION OF PROBLEM

To provide for greater generality of results and to minimize the possibility of overlooking important parameters, the Battelle research program was developed within a systems framework.

The performance of system managers and analysts, as measured in terms of both speed and accuracy, represents the criterion of system performance to be optimized in the present research program. For the system under study, it is apparent that this kind of performance is a function of a variety of variables. The categories of variables believed to be most important are represented in the following general equation:

$$P = f(M, I, D, C, T, O, E),$$

where

P = Operator performance
M = Methods of presentation
I = Type of information displayed
D = Density of information
C = Complexity of information
T = Operator tasks
O = Operator characteristics
E = Environmental factors

These categories of variables are illustrated in Fig. 2, in which a distinction is made between those categories of variables that are under the control of the system designer and those categories that are not. Within the limits of electronic equipment, the system designer has control over methods of presentation. The system designer might also have control over operator characteristics, in so far as he might specify the types of characteristics, such as good visual acuity, high intelligence, etc., necessary for satisfactory performance in the system. Moreover, he controls the environmental factors by specifying the illumination level of the operations room, distance between consoles, etc.

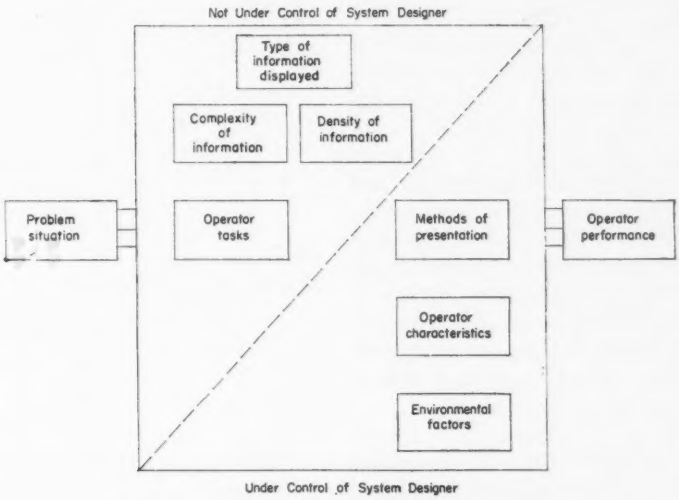


Fig. 2. Variables that affect operator performance.

On the other hand, the system designer has little control over the type of information to be displayed and the operator tasks to be performed, both of which are essentially determined by the particular problem situation that is presented.

The first major step in the present research program was to select for study important variables pertaining to methods of presentation. Almost as important, however, were the

decisions concerning which of the other categories of variables to include in the experiments and which to exclude. Decisions concerning these secondary categories of variables were as follows:

1. Type of information displayed. No attempt would be made to evaluate the effects of various types of information on operator performance. For the information categories to be used in the experiments, however, efforts would be made to develop items expected to be found in the system under study.
2. Density and complexity of information. Because of the likelihood of significant interactions between display conditions and density-complexity of data it was decided to vary both density and complexity of data so that the degree of interaction between these variables and display conditions could be ascertained.
3. Operator tasks. Because of the likelihood of significant interactions between display conditions and operator tasks, it was decided to select representative tasks and, further, to vary them so that the degree of interaction between tasks and display conditions could be ascertained.
4. Operator characteristics. No attempt would be made to determine the relation between operator characteristics and performance. Subjects selected for the experiments, however, should be fairly representative of the population of intelligence specialists assigned to the system under study.
5. Environmental factors. An effort would be made to optimize major environmental conditions and, in so far as possible, to hold these conditions constant throughout the experiment. (Moreover, the size of the displays would be optimized and then held constant.)

The objective of the battelle research program was to develop design criteria for intelligence display formats. Three types of visual displays were to be investigated:

1. Alpha numeric.
2. Trend.
3. Cartographic.

FORMULATION AND SELECTION OF RESEARCH PROBLEMS

After an acceptable conceptual framework for the system had been developed, the next step in the research program was to select

TABLE I

Problem Areas for Alpha-Numeric Displays

Display parameters	*Variables to consider*
(a) Format:	(1) Arrangement of categories
	(2) Use of color
	(3) Black on white versus white on black
	(4) Size of display
	(5) Ambient/display lighting ratio
(b) Density:	(1) Number of categories
	(2) Total number of letters and numbers
	(3) Amount of data per unit area
	(4) "Word" size
(c) Number of Categories:	(1) Data are given for how many categories
(d) Coding Dimensions:	(1) Descriptive phrases versus symbols for categories
	(2) What coding system for symbols
	(3) Families of symbols
	(4) Multidimensional coding
	(5) Use of colored symbols
	(6) Association characteristics
(e) Rate of Change:	(1) Change of categories
	(2) Change of all data for categories
	(3) Change of particular data for categories

specific problems for investigation. Problem areas were formulated by considering display parameters in relation to each of the three types of displays. The results of this formulation are presented in Tables I, II, and III.

As an aid in the selection of problems to be investigated, three primary criteria were established:

(1) Importance of problem.
(2) Applicability of results.
(3) Feasibility of conducting study.

TABLE II

Problem Areas for Trend Displays

Display parameters	*Variables to consider*
(a) Format:	(1) Line graph versus bar graph
	(2) Arrangement of categories
	(3) Number of trends
	(4) Black on white versus white on black
	(5) Size of displays
	(6) Ambient/display lighting ratio
(b) Density:	(1) Number of trend figures
	(2) Amount on each figure
(c) Number of Categories:	(1) Number of graphs
	(2) Number of different lines on graph
(d) Coding Dimensions:	(1) What symbols are used
	(2) How are symbols organized
	(3) Multidimensional coding
	(4) Use of color coding
	(5) Association characteristics
(e) Rate of Change:	(1) Number of time intervals
	(2) Change of present time condition
	(3) Rate of change of new parameter

TABLE III

Problem Areas for Cartographic Displays

Display parameters	*Variables to consider*
(a) Format:	(1) Degree of pictorial reality
	(2) Area covered
	(3) Use of texture
	(4) Use of color
	(5) Use of elevation
	(6) Mosaic or single map
	(7) Black on white versus white on black
	(8) Size of display
	(9) Ambient/display lighting ratio
	(10) Scale
(b) Density:	(1) Amount of detail of geographic nature
	(2) Number of names of cities, etc.
	(3) Number of symbols
(c) Number of Categories:	(1) Types of information
	(2) How many types of symbols
(d) Coding Dimensions:	(1) How are categories coded
	(2) Use of colored symbols
	(3) What dimensions (e.g. size, shape) for each information category
	(4) Families of symbols
	(5) Degree of reality in code-association characteristics
(e) Rate of Change:	(1) Number of different maps per unit time
	(2) Rate of updating symbols on maps

By applying these three criteria for problem selection to the problem areas specified in Tables I, II, and III, five specific problems were selected for investigation. The five resulting studies were:

(1) A comparison of vertical and horizontal arrangements of alpha-numeric material.

(2) An evaluation of formats for trend displays.
(3) An evaluation of methods for presentation of graphic multiple trends.
(4) An evaluation of five different abstract coding methods.
(5) An evaluation of the effect of selected combinations of target and background coding on map-reading performance.

EXPERIMENTAL RATIONALE

As mentioned previously, the primary objective of the research program was to develop design criteria for intelligence display formats. This means, of course, that the experimental results should be formulated in such a manner to be useful to the design engineer.

The major factor that reduces the experimentalist's confidence in generalizing to the operational setting is the obvious gap between the research laboratory and the "real world". To alleviate this problem, the present research program was designed as a partial simulation of the actual operational setting. For example, the major experimental conditions, the secondary variables, the subjects, and the operator tasks were all believed to be fairly representative of what might be found in the present system. No attempt was made, however, to simulate exactly the conditions as found in the actual system.

The present research program thus falls somewhere between fundamental laboratory research at one extreme and operational field testing at the other. Such an approach permits the design of well-controlled experiments in which the variables under study may be varied in a systematic manner; furthermore, the introduction of a certain amount of realism permits a greater degree of generalization. The development of a research program at this level on the "degree-of-reality continuum" means essentially this: If it is concluded that one experimental condition is superior to a second condition, then it is assumed that this same *order* of conditions will prevail in the actual operational system. Because the actual system was not completely simulated, however, we do not have as much confidence that the *magnitude* of the differences between experimental conditions will hold in the actual system.

Experimental Apparatus

Two display units were constructed for rear-projection presentation of 35-mm slides. The projectors were fully automatic with respect to slide changing; a tray of 40 cardboard-mounted 35-mm slides could be shown individually simply be pressing a changing switch each time a new slide was desired. The long axis of a 35-mm slide projected at 6 ft was 30 in.

Fig. 3. A display unit.

A display unit consisted of an enclosed housing built on a frame of 2×4-in. lumber. The frame was 8½ ft long, 39 in. wide, and 25 in. high. Legs at the corners and on each side at the center were mounted on casters, which permitted easy movement of the unit. Including the casters, the overall height of the frame was 28 in. A ½-in. sheet of plywood covered the top of the frame.

A 36×36-in. framed piece of ¼-in. glass Lenscreen[2] was mounted in a vertical position 6 in. from the front of the frame. An area 6 ft

[2] Lenscreen is a high-quality display screen produced by a process of coating plate glass with a thin layer of microscopically small lenses.

in length behind the screen was enclosed with ½-in. plywood. The inside of the enclosed area was painted flat black. A small, circular hole at the rear of the enclosed area permitted projection to the center and rear of the Lenscreen. A small wooden stand at the outside rear of the enclosed area, and mounted on top of the frame, held the projector. A display unit is shown in Fig. 3.

The response-timing and stimulus-presentation devices consisted of two control boxes,

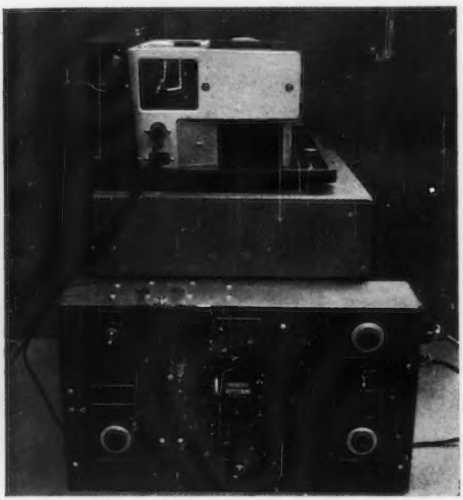


Fig. 4. Rear of display unit showing projector and experimenter's control box for response-timing and stimulus-presentation device.

one for the subject's use and the other for the experimenter's use. The subject's control box contained a push-button which could start and stop the timer, as well as introduce and remove a slide. The experimenter's control box contained a time meter that recorded time with $\frac{1}{10}$ of a second accuracy and a push-button for signaling the subject. The rear of a display unit showing the projector and the experimenter's control box for the response-timing and stimulus-presentation device is shown in Fig. 4.

Regular classroom desk seats were used so that subjects sitting in front of the display unit would have convenient writing space for paper-and-pencil tests. The subject control boxes for the response-timing and stimulus-presentation devices were mounted on the front center, and flush with the desk of each subject's seat.

In all the experiments that employed slides for stimulus material, the above display units were used. In two of the experiments, however, cardboard posters rather than slides were used for stimulus material. The experimental room for these two studies contained a wooden stand with a window blind for the posters and a desk-type chair for the subjects. The illumination was approximately 25 ft.-c. on the screen of the wooden stand, and on the order of 10 to 30 ft.-c. on the Lenscreen.

Selection of Subjects

The population of people expected to be represented by the experimental subjects is indeed a general population, with few specified characteristics. It probably can be assumed, however, that the intelligence specialists working with "second-order" displays in the system under study: (1) are male, (2) are of college caliber in intelligence and technical competence, and (3) possess normal vision.

Thus, to meet these general criteria, 30 male Battelle technicians and engineers were asked to serve as subjects in the experiments. The Snellen Chart and the AO H-R-R Pseudoisochromatic Plates were used to test the subjects for visual acuity and for color vision, respectively.

GENERAL DISCUSSION AND CONCLUSIONS

The present research program has added further support to the value of multivariable experiments. The overall efficiency of such experiments in terms of amount of information obtained per unit cost is indeed sufficiently great to offset any difficulties associated with such an approach. Even more important, however, is that multivariable experiments

permit the investigation of interactions between (and among) variables. It is apparent that design recommendations based on only main effects can be misleading. In most of the experiments, to follow, the results are qualified by the various interactions. Naturally, if such interactions are so small that they have little practical significance—even though they are statistically significant—then the designer need not be burdened with such findings. When such interactions are important in a practical sense, however, it is essential that recommendations be qualified in accordance with these findings.

A second general conclusion concerns the study of operator tasks. First, it is important to note that significant interactions were found between display conditions and operator tasks. Second, it is equally important to note that independent, or orthogonal, task factors were obtained from the statistical analysis. Inasmuch as task tests—rather than task items—were intercorrelated, the resulting factors were not so "clean" as might be desired. Nevertheless, the statistical results permitted meaningful interpretation of the factors. On the basis of these findings, it seems that many of the future studies on the design of visual displays would benefit if operator task were included as one of the experimental variables. Furthermore, because task factors interact with display parameters, it is important to determine as soon as possible whether or not these task factors are representative of the real-world data-processing operations under study. It would then be possible to make recommendations concerning display parameters according to the relative importance of the various tasks.

A review of the literature emphasizes the need for more standardized (as well as more appropriate) conditions for the study of visual displays. It is a relatively easy task to find studies on visual displays that report contradictory results. By reviewing these studies in detail, however, it might be found that they differed in: subject tasks, actual specification of experimental conditions, experimental procedures, and environmental conditions, as well as characteristics of subjects. If greater standardization could be achieved with respect to the important parameters of visual-display experiments, then it is certain that more consistent and more meaningful reports would be forthcoming.

To assist the individuals responsible for the design of visual displays, an extremely worthwhile goal would be to develop a detailed, up-to-date handbook on "Design Criteria for Visual Displays", which might serve as a "Volume II" to the handbook developed by Baker and Grether.[1] Much information is already available for such a handbook, but it is scattered throughout the many Government reports, technical journal articles, and textbooks. If such information were collected and integrated, however, it would serve as an excellent foundation for a visual-display handbook, as well as give direction to future work on visual displays.

REFERENCE

1. Baker, C. A., and Grether, W. F., *Visual Presentation of Information,* WADC Technical Report 54–160, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

A Comparison of Vertical and Horizontal Arrangements of Alpha-Numeric Material—Experiment 1[1]

JOHN L. COFFEY, Battelle Memorial Institute, Columbus, Ohio.

The objective of this study was to determine the relative effectiveness of visual displays containing alpha-numeric material displayed in vertical and horizontal arrangements. Variables included in the experimental design were: types of arrangement of display material, density of material, composition of material, and operator tasks. The major finding in the study was the non-significance of the arrangement variable. It was found that, for all practical purposes, the differential effects of vertical and horizontal arrangement of alpha-numeric materials on operator performance are negligible.

INTRODUCTION

The use of alpha-numeric materials in visual displays is a common practice. These displays, commonly known as status boards, may stand alone or they may be used to complement other types of displays.

The arrangement of the data on the display is one of the basic considerations in the construction of an alpha-numeric display. If there is an optimum way of arranging these data according to the type of material to be used, the amount of material to be used, and the task required of the operator, then the display should be designed to maximize the speed and accuracy with which the material can be used.

The objective of this study was to determine the relative effectiveness of visual displays containing alpha-numeric material displayed in vertical and horizontal arrangements. The relative effectiveness of these two arrangements was evaluated, with variations in the density of the material on the display, the types of alpha-numeric material, and the subject tasks.

METHOD

Experimental Design

The experimental design for the study was a $2 \times 2 \times 4 \times 4$ complete factorial with 12 subjects per cell. Two replications of the study were conducted. The variables in the design were: (1) two types of arrangement—horizontal and vertical; (2) two density levels—one column or row and three columns or rows; (3) four compositions of alpha-numeric material—letters, numbers, letters and numbers combined, and two-letter words; and (4) four subject tasks—counting, locating, identifying, and comparing.

Each subject was tested under each of the 64 experimental conditions in each of the two replications. The same subjects were used in both replications.

Subjects

The 12 subjects for this study were selected from the pool of subjects referred to previously. All subjects were male, with an age range of 22 to 33. The educational range was from high-school graduates with some university undergraduate work to university graduates with varying amounts of graduate work.

Display Material

Sixteen original formats were prepared. Four of these formats contained only letters in different arrangements and density; four formats contained only numbers in different arrangements and density; four formats contained only letters and numbers combined in different arrangements and density; and four formats contained only two-letter words in different arrangements and density.

Letter formats. The four formats containing only letters were constructed in the following

[1] This work was sponsored by the U.S. Air Force under Contract AF 30(602)-2078 (ref. RADC-TR-60-201).

EDUCATION LIBRARY

manner. The first format consisted of one row of 24 letters. The letters were determined by random sampling with replacement from the alphabet of 25 letters (i.e. all letters with the exception of "O"). The second format containing only letters was made up of one column of 24 letters. The third format containing only letters was made up of three rows of 24 letters each. The fourth, and last, format in this group was made up of 3 columns of 24 letters each. In all cases, the sampling was random with replacement.

Number formats. The four formats containing only numbers were constructed in exactly the same manner as those containing only letters. The construction of these four formats resulted in one format containing one row of 24 randomly selected numbers, one format containing one column of 24 randomly selected numbers, one format containing three rows of 24 randomly selected numbers each, and one format containing three columns of 24 randomly selected numbers each (zero was omitted).

Letter and number formats. The four formats containing letters and numbers combined were constructed in essentially the same manner as those for letters alone. In every case, however, each row or column contained 12 numbers and 12 letters. The assignment of the position in the row or column to be occupied by a letter or number was completely randomized. This procedure resulted in one format containing one row of 24 randomly selected letters and numbers, one format containing one column of 24 randomly selected letters and numbers, one format containing three rows each with 24 randomly selected letters and numbers, and one format containing three columns each with 24 randomly selected letters and numbers.

Two-letter word formats. The four formats containing two-letter words were constructed from a population of 16 commonly used words. The formats were constructed in a random manner, such as was done for letters alone or numbers alone. The only difference was that each column or row contained only 16 two-letter words rather than the 24 entries previously used. The construction procedure resulted in one format containing one row of 16 randomly selected two-letter words, one format containing one column of 16 randomly selected two-letter words, one format containing three rows each with 16 randomly selected two-letter words, and 1 format containing three columns each with 16 randomly selected two-letter words.

The 16 formats constructed as outlined above were used for the entire study. These formats contained the variable of horizontal arrangement and vertical arrangement corresponding to rows and columns, the variable of density represented by one row or column and three rows or columns, and the variable of composition of alpha-numeric material represented by letters alone, numbers alone, letters and numbers combined, and two-letter words.

Test Material

The fourth variable in the study, as outlined in the Experimental Design section, was subject tasks. The four tasks employed were counting, locating, identifying, and comparing. For each display format a group of four questions was constructed for each task. The resulting 256 questions were used throughout the study.

Preparation of Experimental Material

The experimental material was prepared by combining appropriate stimulus and test material. In order for each of the tasks to be paired with each display format, the 16 formats were photographed four times each, resulting in a total of 64 slides. Thus, in one replication all subjects viewed each of the original 16 formats at four different times, with a different task to perform each time.

A slide paired with a set of four questions of one type, i.e. count, locate, identify, or compare, constituted one experimental condition. The study was made up of 64 such experimental conditions. Twenty-four random sequences of the 64 experimental conditions were prepared; the only restriction placed upon these sequences was that the same format could not

appear twice in succession. Twelve of these sequences were drawn at random, and the 12 subjects were assigned at random to these sequences. For the second replication of the experiment, the 12 subjects were again assigned at random to the remaining sequences.

For each sequence, a test booklet containing 64 consecutively numbered pages was prepared. Each page contained four questions of one particular type.

The 64 slides were also arranged in randomized sequence to correspond exactly to the

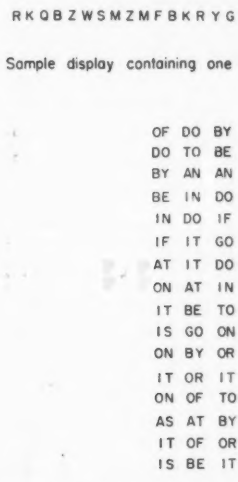

R K Q B Z W S M Z M F B K R Y G T A I W U I S X
Sample display containing one row of letters
OF DO BY
DO TO BE
BY AN AN
BE IN DO
IN DO IF
IF IT GO
AT IT DO
ON AT IN
IT BE TO
IS GO ON
ON BY OR
IT OR IT
ON OF TO
AS AT BY
IT OF OR
IS BE IT
Sample display containing three columns of two-letter words

Fig. 1. Examples of stimulus materials.

arrangement of test-booklet pages. This procedure was maintained for all subjects in both replications of the study. Two sample displays and four sample questions of each type relating to one of the displays are included in Fig. 1 and Table I, respectively.

To permit subject practice and to minimize learning effects, two additional formats were constructed in the previously outlined manner for use as training displays. One format contained one row of 24 randomly selected letters, and the other format contained three columns each consisting of 16 randomly selected two-letter words. For each sample format a set of four questions was developed for each of the four operator tasks. The sample formats were photographed four times to allow pairing of each set of questions with each format. This procedure resulted in eight slides with 32 questions for the training period.

TABLE I

Sample Questions for Sample Display Containing one Row of Letters

Counting Questions

1. How many times does *H* appear?
2. How many times does *K* appear?
3. How many times does *Z* appear?
4. How many times does *E* appear?

Locating Questions

1. Is any *I* adjacent to any *Q*?
2. Is there at least one *Z* anywhere to the right of a *B*?
3. Is there at least one *R* anywhere to the left of an *L*?
4. Is any *H* adjacent to any *K*?

Identifying Questions

1. Is the eighteenth entry from the left an *H*?
2. Is the seventh entry from the left a *Q*?
3. Is the twenty-fourth entry from the right a *C*?
4. Is the twelfth entry from the right a *Y*?

Comparing Questions

1. If the row were divided in the center, would *H* appear in both halves?
2. If the row were divided in the center, would *D* appear in both halves?
3. If the row were divided in the center, would *R* appear in both halves?
4. If the row were divided in the center, would *Y* appear in both halves?

Experimental Procedure

Four experimental sessions were arranged. The first session included the training period and the first 32 experimental conditions. The second session consisted of the final 32 conditions of the first replication. The third and fourth sessions completed the second replication.

The subject's desk chair was placed 5 ft in

front of the screen. All projected alphanumeric characters were approximately ½ in. in height. The center of the projected display was 47 in. from the floor.

During the first part of an experimental session, the subject was read a standard set of instructions. He was told that the purpose of the study was to determine the speed and accuracy with which he could handle alphanumeric material. It was emphasized that speed and accuracy were equally important. After all questions had been answered, the subject was given his appropriate test booklets, and after a signal to commence, the subject proceeded in a self-paced manner to complete the first experimental session.

The second, third, and fourth experimental sessions were conducted in the same manner

TABLE II

Summary of Analysis of Variance

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
Subjects (S)	11	2.1500	Residual	153.57*
Arrangements (A)	1	0.0000	SA	0.00
Density levels (D)	1	11.2709	SD	383.36*
Composition of alphanumeric material (C)	3	0.1958	SC	11.00*
Tasks (T)	3	1.0570	ST	12.42*
SA	11	0.0251	Residual	1.79
SD	11	0.0294	Residual	2.10*
SC	33	0.0178	Residual	1.27
ST	33	0.0851	Residual	6.08*
AD	1	0.0897	SAD	6.45*
AC	3	0.0806	SAC	4.55*
AT	3	0.2455	SAT	23.38*
DC	3	0.0321	SDC	3.49*
DT	3	1.6769	SDT	145.82*
CT	9	0.0404	SCT	2.57*
SAD	11	0.0139	Residual	0.99
SAC	33	0.0177	Residual	1.26
SAT	33	0.0105	Residual	0.75
SDC	33	0.0092	Residual	0.66
SCT	99	0.0157	Residual	0.82
ADC	3	0.1331	SADC	11.18*
ADT	3	0.0331	SADT	2.36
ACT	9	0.0277	SACT	2.50*
DCT	9	0.0291	SDCT	2.39*
SADC	33	0.0119	Residual	0.85
SADT	33	0.0140	Residual	1.00
SACT	99	0.0111	Residual	0.79
SDCT	99	0.0122	Residual	0.87
ADCT	9	0.0240	SADCT	1.68
SADCT	99	0.0143	Residual	1.02
Residual	768	0.0140		

* $p < 0.05$.

as the first, with the exception that training was eliminated.

RESULTS

To give weight to both speed and accuracy of response, the time scores and the accuracy scores were combined into correct-items-per-minute scores. A time score for each page of four questions was recorded during the experiment. The number of correct answers on the page was divided by the time score for the page to obtain a correct-items-per-minute score. The correlation computed between mean subject response times and mean subject correct responses over all experimental conditions was −0.64.

A correlation of 0.93 was found between the two replications of the study. This correlation was based upon mean subject performance for the two replications over the 64 experimental conditions.

In order to more nearly meet the assumptions of homogeneity of variance and normality of distribution for analysis of variance, the correct-items-per-minute scores were converted to logarithmic scores. With the assumption of a mixed model, the 1536 transformed scores were analyzed by analysis of variance techniques. The main effects of density levels, composition of alpha-numeric material, and subject tasks were significant at the 5 per cent level of confidence, while the main effect of arrangement was not. The results of the analysis are presented in Table II.

The significant main effects and two-factor and three-factor interactions from the analysis of variance were analyzed by means of the Duncan multiple range test. All of these analyses were conducted on the means of the logarithmic correct-items-per-minute scores.

The main effect of arrangements considered independently was not significant. The geometric mean (i.e. the antilog of the log mean) of the scores for horizontal arrangement was 6.34 correct responses per minute, and for the vertical arrangement it was also 6.34 correct responses per minute.

The main effect of density levels was statistically significant. As might be expected, performance was superior at the lower density level of one column or row of alpha-numeric material as compared with the higher density level of three columns or rows. The geometric mean of the scores for the lower density level was 7.72 correct responses per minute, and for the higher density level is was 5.20.

The main effect of composition of alpha-numeric material was also significant. The use of two-letter words resulted in better performance than the use of letters alone, numbers alone, and letters and numbers combined. Moreover, the main effect of subject tasks was significant. The tasks of locating and comparing were accomplished significantly better than either the counting or identifying tasks.

The results for all main effects are summarized in Table III.

As indicated from the analysis of variance, a number of the interactions were statistically significant at the 5 per cent level of confidence.

TABLE III

Subject Performance for Each of the Main Effects

(*a*) *Arrangement*

Vertical	*Horizontal*
6.34	6.34

(*b*) *Density*

Three rows or columns	*One row or column*
5.20	7.72

(*c*) *Composition of Material*

Numbers	*Letters*	*Letters and numbers*	*Words*
6.11	6.13	6.30	6.84

(*d*) *Subject tasks*

Count	*Identify*	*Locate*	*Compare*
5.72	5.73	6.75	7.29

Notes:

(1) Scores reported in terms of geometric mean of correct responses per minute.

(2) Experimental conditions connected by line are not significantly different at $p < 0.05$.

The magnitude of these interactions, however, was not sufficiently large to indicate that they would have any meaning in the practical situation. Hence, the interactions will be presented in no greater detail than is shown in Table II.

DISCUSSION AND CONCLUSIONS

This study was designed to compare vertical presentation with horizontal presentation of alpha-numeric material under a variety of conditions. The conclusions are naturally limited to the types and variations of the variables included, to the measures of performance employed in the study, and to the population of subjects represented by those participating in the study.

It was found that subjects performed better with words than with other types of alpha-numeric materials. This probably can be attributed to the greater meaningfulness of words. As would be expected, the higher density condition degraded subject performance. With regard to subject tasks, locating and comparing resulted in significantly better performance scores than did counting and identifying.

The major finding in the study was the non-significance of the arrangement variable. According to the results of the present study, the design engineer can assume that, for all practical purposes, the differential effects of horizontal and vertical arrangement of alpha-numeric material on operator performance are negligible.

An Evaluation of Formats for Graphic Trend Displays—Experiment II[1]

HOWARD G. SCHUTZ, Battelle Memorial Institute, Columbus, Ohio.

This study was designed to determine which of three types of trend formats results in superior performance for a task requiring the subject to make complex decisions. Three commonly used formats were included in the study: line type, vertical-bar type, and horizontal-bar type. Two secondary independent variables were: number of time points and amount of missing data. Results of the study indicate that preference should be given to line-type graphs, followed closely by the vertical-bar type. A secondary finding was that irrelevant points and missing data on graphic trend displays represent important factors in the degradation of operator performance.

INTRODUCTION

The display of trends is usually shown on a graph in which the horizontal axis represents one variable and the vertical axis a second variable. This is especially common when one variable is time and the second is some parameter which will vary in some manner over time. There are many applications for such trend graphs in intelligence systems, because the data collected in such systems often represent values for some parameter at various points in time.

One possible advantage of the graphic trend display over the alpha-numeric type for displaying time-correlated data is its superiority for prediction. Since rate of change can be estimated from the slope of a line, it is possible by extrapolation to estimate when a variable will reach some predetermined critical value. Alpha-numeric displays do not facilitate this type of extrapolation. Another advantage of graphic trend displays is that two or more trends may be readily compared. For instance, the task might be to determine which of two trends is increasing at a faster rate. This also would be a more difficult task with an alpha-numeric display.

In the design of a graphic trend display, one of the first decisions the designer must make concerns the type of format to be used. Format as used with trend displays is defined as the manner in which the data are represented. Three of the most common formats are: (1) the line type which connects data points by a single continuous line; (2) the vertical-bar type which uses vertical lines of varying heights; and (3) the horizontal-bar type which uses horizontal lines of varying length. Although for some applications one type of format may be preferred over the others, the choice is often difficult. Furthermore, the necessary guidelines are not available in the literature.

The present study was designed to yield information to aid design engineers in selecting the type of format for trend displays. Two secondary independent variables were included in this study: number of time points and amount of missing data. These two variables, which will be described in more detail in the next section, were included to allow greater generality of the format results, as well as to determine their effect on operator performance.

METHOD

Experimental Design

The design was a $3 \times 3 \times 3 \times 10$ complete factorial with two replications. The first variable was format, which had three qualitative levels: line, column, and bar. The second variable was number of points, which had three levels: 6, 12, and 18 points. The third variable was amount of missing data, and this also had three levels: none, one-sixth, and one-third missing. The fourth variable was ten randomly selected subjects. This design

[1] This work was sponsored by the U.S. Air Force under Contract AF30 (602)–2078 (ref. RADC–TR–60–201).

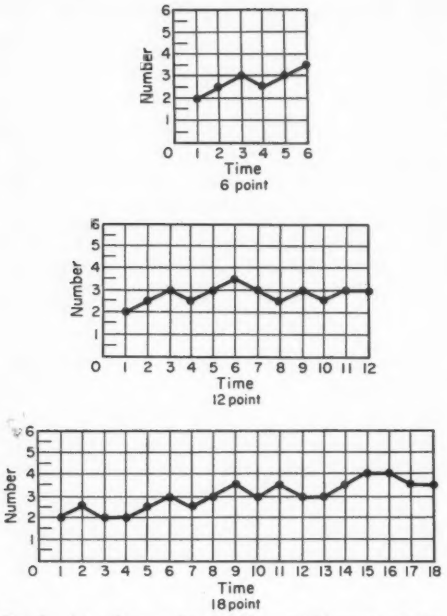


Fig. 1. Line format for 6-, 12-, and 18-point graphs.

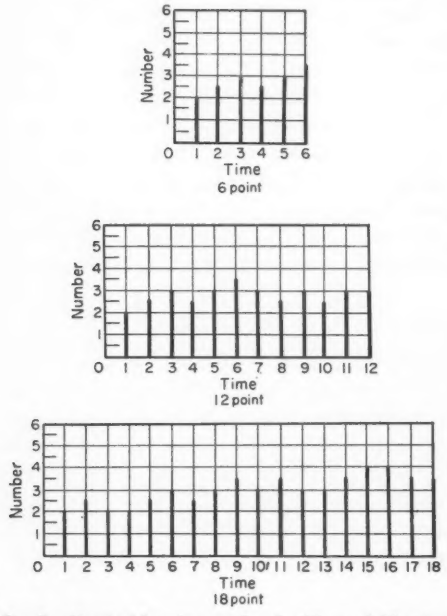


Fig. 2. Vertical-bar format for 6-, 12-, and 18-point graphs.

yielded twenty-seven fixed experimental conditions which, when repeated twice for each subject, resulted in two replications. Fig. 1, 2, and 3_1 show an example of the three formats with 6, 12, and 18 points and no missing data.

The task of the subjects in using these displays involved the manipulation of arbitrary rules to arrive at probabilities. Although somewhat artificial, hopefully, this kind of task may be generalized to the problem-solving

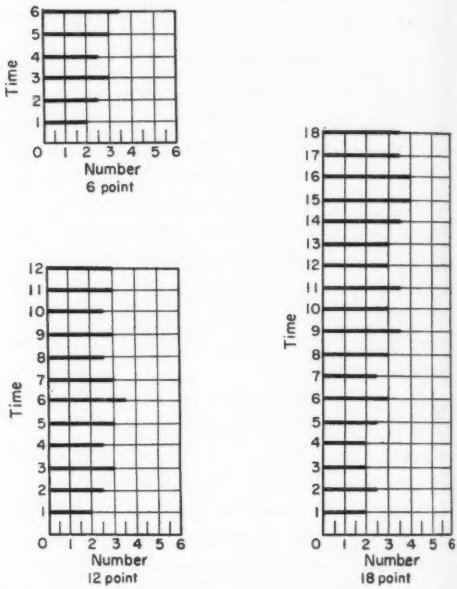


Fig. 3. Horizontal-bar format for 6-, 12-, and 18-point graphs.

behavior of analysts. Since one of the basic kinds of decisions to be made about a graph is whether the trend level is going upward, downward, or remaining essentially the same, this type of decision was selected as the foundation for the operator tasks.

To reduce the effects of individual differences in the interpretation of trends, a set of specific decision rules was taught to each

[1] Stimulus figures are reshown from a print of the original 5 mm slide material and are reasonable facsimiles rather than exact representations.

subject prior to experimentation. These rules correspond to the data displayed in the graphs. Six points were chosen as the basic set of points about which the subject would be required to make decisions. Ten variations of these six points were developed along with corresponding direction and probability of trend values. The probability values represented estimates of the degree of confidence a person might have that the ten types of trends actually exist. In addition, probabilities were modified by the occurrence of missing data in the trends. This modification of the probability was considered to be a necessary change in degree of certainty, which would surely result if there were missing data in a trend. The twelve rules and their corresponding graphic representation for the line-type format with six points are shown in Fig. 4. The four general rules given the subject were as follows:

(1) Points on the graph can differ by only 0.50 of a scale value on the number scale from the point preceding or following it.

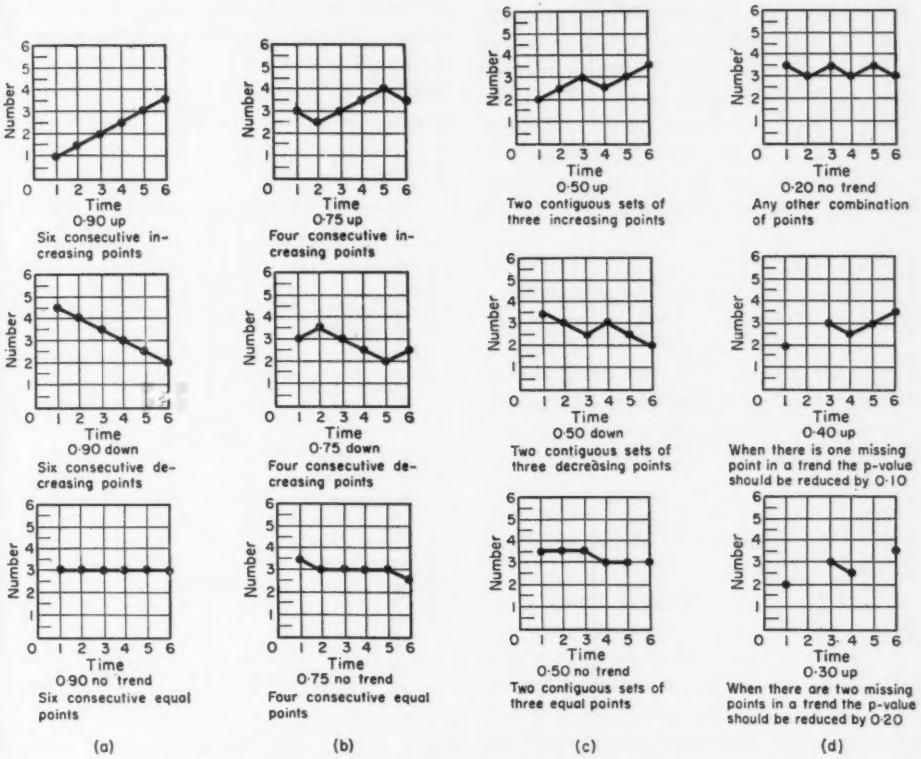


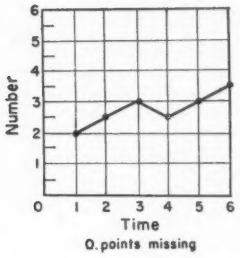
Fig. 4. Twelve patterns and rules for naming trend and determining probability.

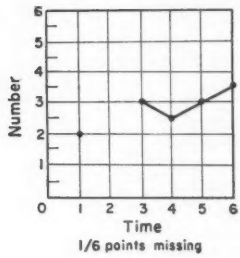
(2) When points are missing, they will be interpolated so as to form a trend according to the twelve specific learned rules and in accordance with the first general rule.

(3) Find the trend with the highest probability of occurrence.

(4) There is only one correct trend for each graph.

Certain conventions were followed in combining the fixed variables. These conventions were developed to create as much realism in a graphic trend display as possible, while still maintaining good experimental control over the various conditions. For the 6-, 12-, and 18-point conditions, the basic set of 6 points was included as a unit. The location of the 6 points was selected at random with the restriction that the 6 points appear as a unit. The additional 6 or 12 points were placed so that they did not satisfy a given rule.





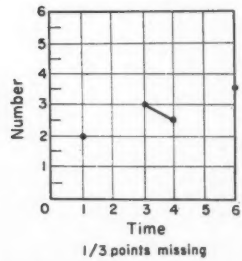
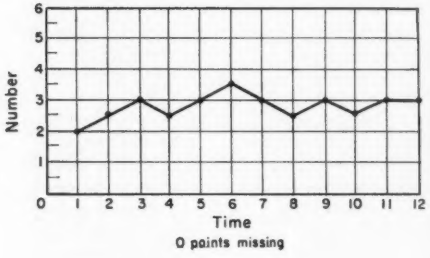
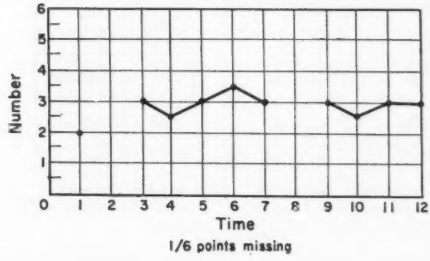


Fig. 5. Examples of 6-point graph with 0, $\frac{1}{6}$, and $\frac{1}{3}$ missing data for line-type graph.

Adjacent points were never more than 0.50 scale points apart, either in the basic set of 6 or in the additional points. For the one-sixth and one-third missing-data conditions, 1 and 2, 2 and 4, and 3 and 6 points were missing from the 6-, 12-, and 18-point conditions, respectively. In the 12- and 18-point conditions, half the points were taken from the basic set of 6 and the other half at random from the remaining points. Two consecutive points were never missing. Fig. 5, 6, and 7 show an example of the line-type format for 6-, 12-, and 18-point graphs which have the same basic set of 6





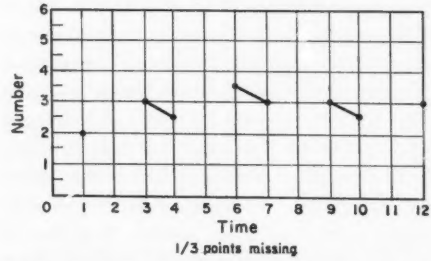


Fig. 6. Examples of 12-point graph with 0, $\frac{1}{6}$, and $\frac{1}{3}$ missing data for line-type graph.

points for the none missing, the one-sixth, and one-third missing point conditions.

A specific design was developed in which each of the twenty-seven treatments would be tested an equal number of times under each of the ten basic patterns of points. In addition, the order of presentation was randomized to minimize effects of learning and fatigue.

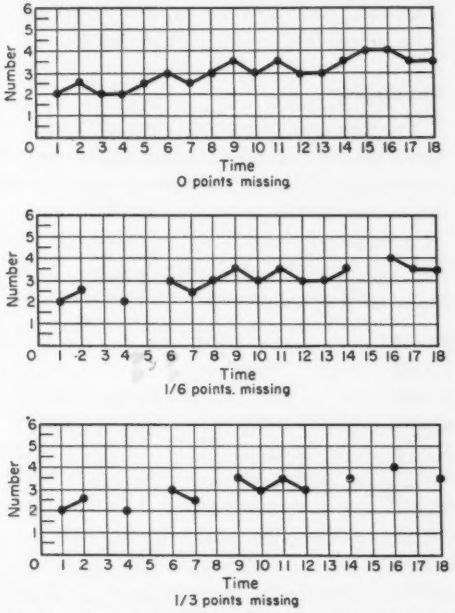


Fig. 7. Examples of 18-point graph with 0, $\frac{1}{6}$, and $\frac{1}{3}$ missing data for line-type graph.

A general comment about the construction of the three types of formats is in order. Every attempt was made to apply sound human-engineering principles to each of the formats. Size of numbers, size of letters, line spacing, and line width were selected to create the best possible representation of the three formats. Human-engineering guides and consultation with technical artists were utilized in making such decisions.

Subjects

Ten male Battelle employees at the professional level were selected from the available pool of subjects. The professional level was considered appropriate since this type of display would most likely be used by officers with considerable experience in intelligence work.

Materials

The twenty-seven display conditions were drawn up by artists and photographed to produce a positive 35-mm transparency. The projection and response-timing equipment were the same as described in the first paper of this series. The projected size of the graphs naturally varied with the variables of format and number of points: 6-, 12-, and 18-point graphs were $5\frac{3}{4}$ in. $\times 5\frac{3}{4}$ in., $5\frac{3}{4}$ in. $\times 11\frac{1}{4}$ in., and $5\frac{3}{8}$ in. $\times 16\frac{3}{4}$ in., respectively.

Procedure

All subjects were given a general orientation on the purpose of the study and the procedures to be followed. They were then instructed to memorize the set of rules shown in Fig. 4. After a subject had learned the list to a 100 per cent criterion, he was given a complete set of twenty-seven treatments, during which time he could ask questions to clarify any uncertainty. This extensive practice session was considered necessary because of the difficulty of the tasks and the necessity of associating the rules with the graphs. The subjects were instructed to find the correct trend as quickly as possible. The testing procedure was self-paced. The subject would depress a push-button to present a display on the screen and to start a timer. As soon as he made a decision, he would depress the push-button again. This would remove the display and stop the timer. The subject would then state the direction and *p*-value associated with the trend.

The subjects on two different occasions received two other sets of the twenty-seven experimental conditions. The procedures were the same as in the training session. The subjects were tested on the twelve rules immediately before each replication to check on their mastery of the material. Each test session required approximately 30 min for completion.

The experimenter recorded both the time to respond to each display and the subject's oral response for the direction and p-value of the trend. Seating and lighting arrangements were identical to those described in the first paper of this series.[2]

After the experiment was completed, each subject ranked the trend formats according to the extent to which he preferred to use the displays.

RESULTS

Treatment of the Raw Data

Means of the time scores over ten subjects were computed for each of the twenty-seven treatments and for both replications. The trend and p-values were scored for accuracy according to the following rules:

Subject's response	*Score*
Wrong direction—wrong probability	0
Right direction—wrong probability	1
Right probability—wrong direction	1
Right direction—right probability	2

Means of the accuracy scores over ten subjects were then computed for each of the twenty-seven treatments and for each replication.

Reliability

The reliability of both the time and accuracy scores were checked by correlating the twenty-seven means for time and accuracy on the first

TABLE I

Summary of Analysis of Variance for Time Scores

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
Subjects (S)	9	1.2162	Residual	44.55*
Format (F)	2	0.6469	SF	23.18*
Number of points (N)	2	4.9481	SN	213.12*
Missing data (M)	2	3.5270	SM	109.52*
Replication (R)	1	0.6112	SR	8.64*
SF	18	0.0279	Residual	1.02
SN	18	0.0232	Residual	—
SM	18	0.0322	Residual	1.18
SR	9	0.0707	Residual	2.59*
FN	4	0.0357	SFN	1.05
FM	4	0.1082	SFM	3.79*
FR	2	0.0154	SFR	—
NM	4	0.0414	SNM	1.64
NR	2	0.0424	SNR	2.10
MR	2	0.0166	SMR	—
SFN	36	0.0340	Residual	1.24
SFM	36	0.0285	Residual	1.04
SNM	36	0.0252	Residual	—
FNM	8	0.0373	SFNM	1.25
SFR	18	0.0384	Residual	1.41
SNR	18	0.0202	Residual	—
SMR	18	0.0323	Residual	1.18
SFNM	72	0.0297	Residual	1.09
Residual	200	0.0273		

* $p < 0.05$.

TABLE II

Geometric Means for Significant Fixed Main Effects and Interactions for Time Scores

Format			*Number of points*			*Missing data*			*Replication*		*Format by missing data*			
Line	*Vertical bar*	*Horizontal bar*	6	12	18	0	$\frac{1}{6}$	$\frac{1}{3}$	1	2	*Proportion missing*	*Line*	*Vertical bar*	*Horizontal bar*
6.81	7.36	8.91	5.15	7.87	11.02	5.50	7.76	10.47	8.26	7.07	0	4.52	5.29	6.94
											$\frac{1}{6}$	6.77	7.39	9.35
											$\frac{1}{3}$	10.34	10.00	10.89

Notes:
(1) All values not underlined by the same line differ from one another at $p < 0.05$.
(2) Scores reported in seconds.

replication with their respective means for the second replication. The r for time was 0.93 and the r for accuracy was 0.75.

The relation between time and accuracy was determined by correlating the time and accuracy averaged over both replications. This correlation was -0.90, which indicates that, for all practical purposes, time and accuracy were reflecting the same pattern of scores over the twenty-seven treatments.

Time-Score Analysis

The raw time scores were transformed to logs in order to more nearly meet the assumptions underlying analysis of variance. In addition, the logs would yield a more valid measure of central tendency, the geometric mean, than would an arithmetic mean (because of the characteristic positively skewed nature of time scores). The transformed time scores were analyzed with a modified five-way mixed model of the analysis of variance. Mean squares were computed for only the main effects and all two- and three-way interactions for the replication variable. These results are summarized in Table I.

All main effects were significant, and two of the two-way interactions, subjects by replication and format by missing data, were significant. Table II gives the geometric means[2] for each of the major effects. A Duncan multiple range test[1] was conducted for each significant main-effect source to determine which differences between pairs of means were significant. All the means for the main effects differed from one another at $p < 0.05$.

Table II also includes the means for each format under all three missing-data conditions, since this interaction was statistically significant. The Duncan multiple range test was also applied to these interaction means. Fig. 8 graphically illustrates the interaction. It is apparent that the three formats do not differ significantly from one another at the one-third level of missing data, whereas at the other two levels of missing data, the formats are significantly different. The significant subject by

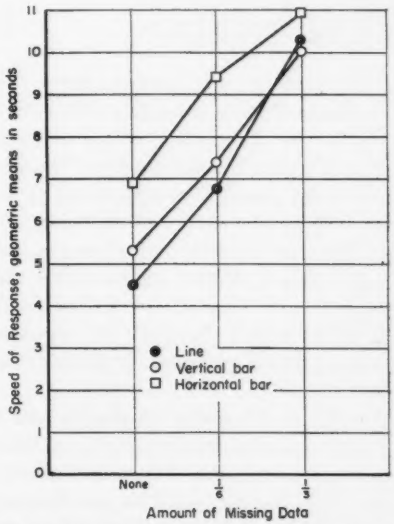


Fig. 8. Graphic illustration of two-way interaction between format and amount of missing data.

[2] Geometric means are antilogs of mean log scores.

replication interaction signifies that some subjects improved in performance for the second replication and others did not.

Accuracy-Scores Analysis

The accuracy scores were not distributed in a manner to meet the assumptions of analysis of variance. Because of the restricted range of scores, transformation of scores would still not provide a set of scores which could be correctly analyzed by a parametric analysis of variance. Thus, because the accuracy scores did not lend themselves to ordinary analysis-of-variance techniques, and, further, because the correlation ($r = -0.90$) between time and accuracy was so high, the accuracy data were not extensively analyzed.

test for this analysis may be attributed in part to the small number of cases available for the analysis. Examination of the means in Table III indicates that the rank order of the means is about the same as in the time-score analysis. This similarity would be expected on the basis of the high r between time and accuracy data. It might be of some additional value to point out that most of the errors occurred in the missing-data conditions, as is evident by examination of the means in Table III.

The preference ranking data were averaged over the ten subjects. The mean ranks were as follows: line, *1.1*; column, *1.9*; and bar, *3.0*, with 1 representing the most preferred graph. This rank order is the same as that found for both the time and the accuracy scores.

TABLE III

Means of Accuracy Scores for Main Variables and One Interaction

*Formal**			*Number of points**			*Missing data**			*Replication*		*Format by missing data*			
Line	*Vertical bar*	*Horizontal bar*	6	12	18	0	$\frac{1}{6}$	$\frac{1}{3}$	1	2	*Proportion missing*	*Line*	*Vertical bar*	*Horizontal bar*
1.72†	1.64	1.50	1.78	1.60	1.48	1.86	1.67	1.33	1.62	1.61	0	1.92	1.95	1.70
											$\frac{1}{6}$	1.80	1.70	1.52
											$\frac{1}{3}$	1.43	1.27	1.28

* Indicates a significant chi square was found in a Friedman two-way analysis of variance.

† 0—wrong direction—wrong probability.
1—right direction—wrong probability.
1—right probability—wrong direction.
2—right direction—right probability.

A Friedman two-way, non-parametric analysis of variance[3] was conducted on the accuracy scores for the three main effects of format, number of points, and missing data. The means for the nine treatment combinations in each were used as the source of variation. The means for these variables are shown in Table III. The almost identical means for the replications did not require an analysis. This analysis, however, resulted in significant differences among the levels for all the remaining variables.

In order to determine whether or not the same interaction existed for accuracy as did for time scores, a two-way Friedman analysis of variance was also conducted for each format by missing-data level. No significant effects were found. The poor discriminating power of this

DISCUSSION AND CONCLUSIONS

The significant effects of format on both speed of performance and accuracy of performance indicate clearly that the line graph resulted in superior performance compared with the vertical-bar or horizontal-bar type. The superiority of the line and vertical-bar graphs over the horizontal bar may be partly attributed to the fact that the two axes are reversed for the horizontal-bar graph; that is, time is on the vertical axis and number is on the horizontal axis.

The interaction between format and missing data indicates that at the higher level of missing data (one-third missing), the three formats do not differ significantly. This is probably due to the fact that the task is so difficult that the differences in format are relatively unimportant.

The number-of-points variable can be viewed as a type of irrelevant data situation. The subject must search for a particular set of four to six points in a context of points which are of no value. The degradation of performance that results from an increase in irrelevant points indicates that for situations comparable to these experimental conditions it would be better to display only the minimum number of points that will be needed in looking for a trend. The effects of missing data on performance would lead one to attempt to supply missing data if at all possible in a graphic trend. The significance of replication in the analysis of variance for time scores demonstrates that performance improved from the first to the second replication. The subject by replication interaction reveals that this change in performance is greater for some subjects than for others, Accuracy scores, however, did not reflect any significant learning from one replication to the other.

The major conclusion of this study is:

> Line-type graphs should be used to show time-correlated data. If a second choice is to be made, the vertical-bar type is recommended.

REFERENCES

1. DUNCAN, D. B., Multiple Range and Multiple F Tests. Reprint Series No. 49; reprinted from *Biometrica*, **11** (1), 1955.
2. HITT, W. D., *et al.*, Development of Design Criteria for Intelligence Display Formats (present edition of this journal).
3. SIEGEL, S., *Nonparametric Statistics*, McGraw-Hill, New York. 1956.

An Evaluation of Methods for Presentation of Graphic Multiple Trends—Experiment III[1]

HOWARD G. SCHUTZ, Battelle Memorial Institute, Columbus, Ohio.

The primary objective of this study was to determine the effect of multiple-line versus multiple-graph presentation of trend-type displays on operator performance. Four types of lines having low confusability were determined experimentally from a sample of twenty-five lines. The primary variable was single-graph, multiple-line presentation versus multiple-graph single-line presentation. Other variables included in this study were: number of lines, degree of confusion among lines, coding of lines, and two operator tasks: point-reading and comparing. It was found that for the point-reading task, either type of display is acceptable, but for the comparing task, the multiple-line display is much superior to the multiple-graph display. Moreover, the use of color coding for the graph lines tended to improve performance slightly.

INTRODUCTION

The third paper in this series[5] led to the conclusion that the line graph is preferred for displaying single-trend data. The present experiment is concerned with methods of presenting multiple trends. Two major alternatives exist: (1) display several trend lines on a single graph, or (2) display single trend lines on several graphs. The existing human-engineering literature does not contain direct comparisons of these two display methods.

As the number of trend lines is increased, the multiple-line display might result in confusion among lines. This kind of confusion would not be found in the multiple-graph situation. On the other hand, making comparisons among lines might prove to be more difficult in the multiple-graph situation because of the larger separation between lines. From the standpoint of space, the multiple-line presentation is obviously superior, but the price paid in performance may be high. With such questions in mind, an experiment was designed to compare multiple-line with multiple-graph presentations. Four secondary independent variables were included in the study to achieve greater generality as well as to determine their own effects on operator performance. These secondary variables were: (1) number of trend lines displayed, (2) confusability of trend lines, (3) color coding versus black and white, and (4) graph-reading tasks.

A preliminary experiment was conducted to obtain line codes that would help keep lines distinct on the multiple-line displays. This preliminary study will be reported first, and then the multiple-line versus multiple-graph experiment will be presented.

PRELIMINARY EXPERIMENT: EVALUATION OF CODES FOR BLACK AND WHITE LINES

Many varieties of line codes are currently used on graphs. (Each experimenter apparently has his favorite line codes for graph construction.) Some lines are easily confused with each other, while others are quite discriminable. This preliminary study was conducted for the purpose of obtaining a set of minimally confused lines for the multiple-line, multiple-graph study. An assumption was made that the degree to which lines would be confused when presented singly for short time intervals under poor viewing conditions would be comparable with the confusion resulting from a high degree of crossing among a set of lines on a graph.

[1] This work was sponsored by the U.S. Air Force under Contract AF 30(602)–2078 (ref. RADC–TR–60–201).

Method

Selection and preparation of line styles. Twenty-five lines varying in style were selected from among the almost limitless variety found in use. Preference was given to lines with relatively simple construction which would be more amenable to presentation by electronic equipment, such as might be used in an intelligence system. Each line was drawn by an artist and photographed on film as a 35-mm slide. This was done twice to yield a total of fifty slides, two for each line. Examples of the lines selected are shown in Fig. 1. The projected line length was 19 in. and the width was $\frac{1}{16}$ in.

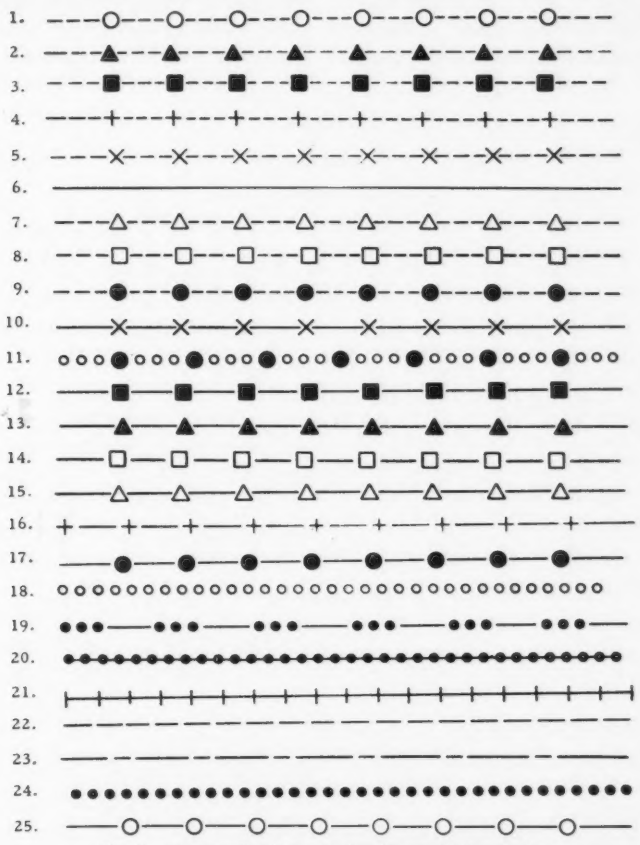


Fig. 1. Twenty-five lines used in line confusion study.

Materials. The experimental materials consisted of the fifty 35-mm slides described above, a rear-projection facility and a subject response box with twenty-five push-buttons. The experimenter observed the subject's responses with a similar box containing lights corresponding to the buttons of the subject's response box.

Subjects. Ten subjects of professional level were drawn from the available pool of subjects.

Experimental Procedure. Each subject was given the twenty-five lines twice in random order. The exposure time was 0.2 sec. The image was blurred by defocusing in order to produce a degree of confusion comparable with that resulting from viewing a single

EDUCATION LIBRARY

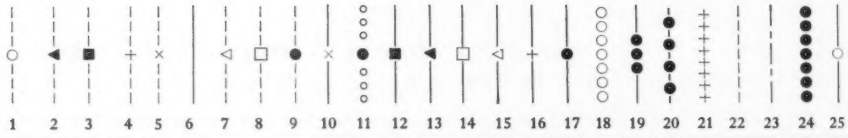
graph with many crossing lines. The subjects were instructed to identify the line shown on the screen by depressing the appropriate response button. They were told that any line could appear any number of times. The sessions were self-paced and lasted approximately 25 min. Each subject was presented all fifty slides during an experimental session.

Analysis of the data and results. The first step in the data analysis was the construction of a confusion matrix. The twenty-five stimulus lines and the responses made to them are presented in Table I. This table shows what proportion of the time each stimulus was identified as one of the twenty-five possible stimuli. The stimulus lines presented are found across the top of the table and the possible responses are presented along the side. The values in the cells represent the proportion of times a particular response was

TABLE I

Confusion Matrix

Stimulus



Response	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	15							30						20	05	10					05				15
2		20		05	10		05		05	05			10			05					05				05
3		05		05	05		05			10	05	10			05		10			05					
4				05												05		05				05		05	
5	05			05	15		05			20						05					05		05		05
6						95								05								25	10		
7	15			05	05		25	05					05	05	20					05					10
8	05			10				10		05				15	05	05									
9	05	20	20	15	15		10		10	15	60	20	30			05	25				05				05
10	05			05							05					05									05
11											15							05			05				
12		05	15	05	05				05	05		15			05	05	05		05		05				
13		05			05		05	05		05		05	20		10	05							05		
14														05		05						05			10
15	05						25	05		05				05	15	05							05		
16	05	05		20	10		10			10	05		05		10	20	10				10			05	05
17	05	25	50	10	20		10	05	65	20		45	20		05	15	45				20				
18				05										05				10			05			05	
19		05			10														95						
20	10	10	10						15		05	05	10		10	05	05	45		50	35			40	10
21				05														15		05			05	05	
22																						45	40	05	
23						05																20	30		
24			05								05							20		35				35	
25	25							40						40	10										30

The proportion of total stimulus presentations resulting in each response. Decimal point omitted.

made to a given stimulus. Numbers in the diagonal represent the proportion of times a stimulus was identified correctly.

Using the constant-ratio rule,[1] this matrix was examined to find possible submatrices of four lines which would result in the least amount of confusion. A submatrix was found which had no confusions among the four lines of which it was composed. In other words, no one of these lines was ever identified as one of the other lines. This submatrix was made up of the following lines:

(No. 4) + − − − + − − − ;
(No. 1) − − − O − − − ;
(No. 13) ——▲——▲——; and
(No. 11) o o o ● o o o ●.

The assumption is made that if only these four lines were presented, the results would have been 100 per cent correct identifications.

Conclusions

Four-line styles were selected which were considered a maximally discriminable set of line styles. These styles, or codes, do not represent the only such set of lines, of course, but they do meet the needs of the major experiment on multiple-line, multiple-graph display techniques.

MAJOR EXPERIMENT: MULTIPLE-TREND DISPLAYS

Experimental design

The design was a $2 \times 3 \times 2 \times 2 \times 2 \times 10$ complete factorial with two replications. The first variable was type of display; that is, multiple line or multiple graph. The second variable was number of lines, which had three levels: two, three, and four lines. The third variable was degree of confusion among lines and had two levels: low confusion and high confusion. The fourth variable was a task variable and consisted of two types of tasks: point reading and comparing. The fifth variable was type of line coding: color or black and white lines. Ten subjects constituted the sixth variable.

The multiple-line and multiple-graph displays included the same data. In the multiple-line displays, two, three, or four lines were presented on a single graph. For the multiple-graph displays, the same lines were displayed in a vertical array on two, three, or four graphs—one line on each graph. Fig. 2 shows the same data presented by both methods.

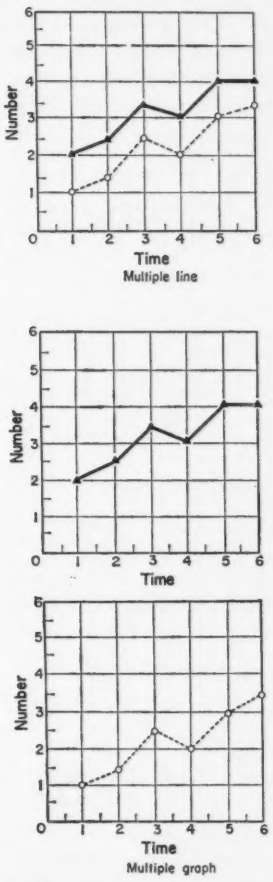


Fig. 2. Examples of multiple-line and multiple-graph presentations.

The use of two, three, and four lines in the

[1] Stimulus figures are reshown from a print of the original 3-mm slide material and are reasonable facsimiles rather than exact representations.

experiment resulted from a desire to make the results more general than if only two lines had been used. Also, differences between the two types of presentation might depend on the number of lines or graphs involved.

Degree of confusion was also varied in the experiment. For low confusion, the lines never crossed. For high confusion, the lines crossed four, seven, and ten times for the 2-, 3-, and 4-line conditions, respectively. Fig. 3 shows an

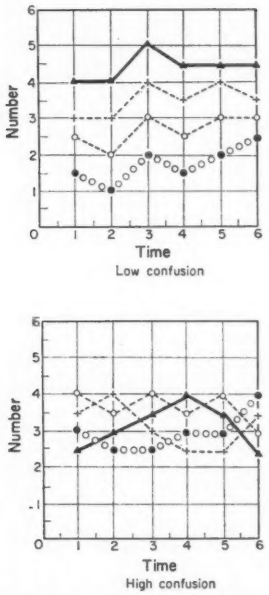


Fig. 3. Example of low and high confusion among lines.

example of low and high confusion conditions for four lines in a multiple-line display.

Two types of coding, color and black and white, were used in the experiment. The use of color as an aid to visual displays is often recommended, but sound experimental evidence concerning its contribution is not available. The four black and white codes used in the present experiment were selected from the preliminary study. Four colored lines—red, yellow, green, and purple—were selected from a list of recommended colors given by Conover and Kraft.[2]

The experiment included two operator tasks: point reading and comparing. These tasks represent two basic operations that are important in most graph applications. The point-reading task consisted of reading vertical-axis values from specified horizontal-axis values for a given line. The comparing task consisted of determining the line with the highest value on the vertical axis for specified values on the horizontal axis.

These five variables resulted in forty-eight fixed experimental conditions. These conditions were presented to ten subjects on two different occasions.

Method

Display. To represent the forty-eight fixed conditions, six basic sets of points on the graph were constructed. The horizontal axis represented time and the vertical axis represented number. All experimental conditions contained the same set of six points in the appropriate form for that experimental condition. Fig. 4 shows all six basic sets of displays for the multiple-line and black-and-white codes. Grid lines were placed at each unit division on both the horizontal axis and vertical axis as an aid in reading the numerical values. Small pips marked the halfway points between units as an additional aid. The forty-eight drawings were photographed on Super Anscochrome 35-mm film which yielded positive transparencies. Four practice slides were constructed: (1) multiple-line, two-line, black-and-white; (2) same as the first but in color; (3) multiple-graph, two-line, black-and-white; and (4) same as the third but in color. The projected sizes of the displays were as follows: single graph—$5\frac{1}{4}$ in. $\times$ $5\frac{1}{4}$ in.; two graphs—$12\frac{1}{4}$ in. $\times$ $5\frac{1}{4}$ in.; three graphs—$19\frac{1}{4}$ in. $\times$ $5\frac{1}{4}$ in.; four graphs—$26\frac{1}{4}$ in. $\times$ $5\frac{1}{4}$ in.

Certain conventions were followed in combining the number of lines and code variables. The two-line codes were composed of O– – – and ——▲——. The three-line codes used

O– – –, ——▲——, and + – – – + – – –. The four-line codes used all four styles. For the color condition, red and green were used for two-line codes; red, green, and yellow were used for three-line codes.

Test booklets. Twenty test booklets were assembled, ten for each replication. Each booklet consisted of fifty-two pages. The first four pages were used with the practice slides and the next forty-eight with the experimental conditions. The four practice slides were in the same order for all subjects, but the order was random for the forty-eight experimental conditions.

To obtain a sample of performance for the tasks, on each page of the test booklet there appeared five time values, for which the subject was instructed to read off the numerical value of a given line or else specify which line was highest, depending on whether it was a point-reading or comparing task. These five values were chosen from a set of ten randomly selected time values from the eleven possible half-interval units. The ten sets of five questions were assigned to the forty-eight experimental conditions in such a manner that each condition would be tested an equal number of times for a given set of five time values. The lines utilized for the point-teading task were assigned in a random manner.

Equipment. The projector and response-timing system were those described in the first paper in this series.[4]

Subjects. Ten male Battelle employees at the professional level were selected from the pool of available subjects.

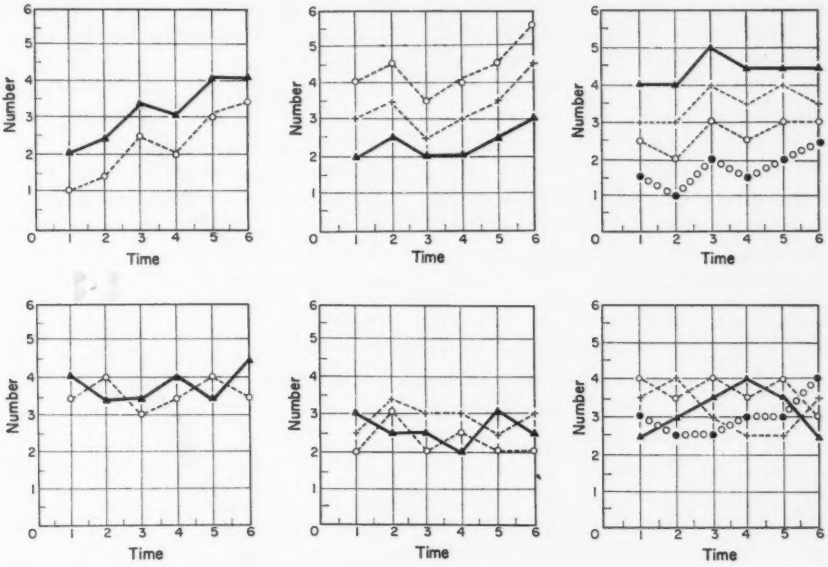


Fig. 4. The six basic sets of points used for all experimental conditions.

Training. The training consisted of an orientation period in which the subjects were told the purpose of the study, shown the types of presentation, and given specific instructions as to the procedures to be used in the experiment. Only a minimum amount of training was considered necessary, inasmuch as the task was one with which the subjects were already familiar.

Experimental procedure. The ten subjects were tested individually on two different occasions,

once for each replication. Each subject was first given the four practice slides, then the forty-eight experimental slides. The subject was given a task booklet selected at random and instructed to read the first page to determine what line he was to utilize and whether he was working on a point-reading or a comparing task. The subject would then depress a push-button which brought up the first slide. After answering the five questions for that slide, he was instructed to depress the push-button again, removing the slide. The experimenter then recorded the time elapsed since the slide appeared before the subject and signalled the subject to proceed to the next page of the booklet. Thus, the progress of the test session was essentially self-paced. This procedure was followed for the remaining three practice slides and for the forty-eight experimental slides. Each experimental session lasted approximately 30 min. Seating and lighting arrangements were identical to those described in the first paper in this series.[4]

After the experiment was completed, each subject rank ordered the display conditions to indicate how much he liked using them.

Results

Treatment of the data. The time scores were organized by the forty-eight treatments, and means over ten subjects were computed for each replication. The accuracy scores (number right) were combined in the same manner. There were very few errors (only 64 out of 960, or 1.33 per cent), and the errors that did occur were not concentrated in any one treatment. Therefore, no further analysis of the accuracy data was necessary.

Reliability. To determine the reliability of measurement for this experiment, the means computed for the ten subjects for the forty-eight treatments on the first replication were correlated with the corresponding data for the second replication. This correlation was 0.98, indicating that the pattern of means for the first and second replications was essentially the same.

TABLE II

Summary of Analysis of Variance for Time Scores

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
Subjects (S)	9	0.6410	Replication	75.54*
Display (D)	1	1.6769	SD	88.72*
Number of lines (N)	2	0.0616	SN	11.94*
Confusion (C)	1	0.5246	SC	111.88*
Task (T)	1	2.0922	ST	840.58*
Line code (L)	1	0.1081	SL	17.56*
DN	2	0.1299	SDN	15.48*
DC	1	0.0013	SDC	—
DT	1	1.6644	SDT	195.05*
DL	1	0.0206	SDL	2.98
SD	9	0.0189	Replication	2.23*
NC	2	0.0830	SNC	14.13*
NT	2	0.0504	SNT	10.06*
NL	2	0.0132	SNL	2.35
SN	18	0.0052	Replication	—
CT	1	0.1566	SCT	21.62*
CL	1	0.0167	SCL	2.36
SC	9	0.0047	Replication	—
TL	1	0.0294	STL	4.45

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
ST	9	0.0025	Replication	—
SL	9	0.0062	Replication	—
DNC	2	0.0024	SDNC	—
DNT	2	0.0599	SDNT	11.04*
DNL	2	0.0010	SDNL	—
SDN	18	0.0084	Replication	—
DCT	1	0.1182	SDCT	23.23*
DCL	1	0.0032	SDCL	—
SDC	9	0.0089	Replication	1.04
DTL	1	0.0025	SDTL	—
SDT	9	0.0085	Replication	1.01
SDL	9	0.0069	Replication	—
NCT	2	0.0233	SNCT	4.46*
NCL	2	0.0192	SNCL	3.27
SNC	18	0.0059	Replication	—
NTL	2	0.0098	SNTL	1.36
SNT	18	0.0050	Replication	—
SNL	18	0.0056	Replication	—
CTL	1	0.0027	SCTL	—
SCT	9	0.0072	Replication	—
SCL	9	0.0071	Replication	—
STL	9	0.0066	Replication	—
DNCT	2	0.0300	SDNCT	5.43*
DNCL	2	0.0292	SDNCL	3.20
SDNC	18	0.0033	Replication	—
DNTL	2	0.0064	SDNTL	—
SDNT	18	0.0054	Replication	—
SDNL	18	0.0052	Replication	—
SDCL	9	0.0035	Replication	—
SDCT	9	0.0051	Replication	—
DCTL	1	0.0185	SDCTL	5.34*
SDTL	9	0.0061	Replication	—
NCTL	2	0.0040	SNCTL	—
SNCT	18	0.0052	Replication	—
SNCL	18	0.0059	Replication	—
SNTL	18	0.0075	Replication	—
SCTL	9	0.0055	Replication	—
DNCTL	2	0.0034	SDNCTL	—
SDNCT	18	0.0055	Replication	—
SDCTL	9	0.0035	Replication	—
SDNTL	18	0.0091	Replication	1.07
SDNCL	18	0.0091	Replication	1.07
SNCTL	18	0.0075	Replication	—
SDNCTL	18	0.0113	Replication	1.33
Replication	480	0.0084		

* $p < 0.05$

Analysis of variance. The raw time scores were converted to logarithms in order to more nearly meet the assumptions underlying analysis of variance. In addition, because of the characteristic positively skewed distribution of time scores, logs yield a more valid measure of central tendency—the geometric mean—than would an arithmetic mean.

A six-way mixed model of the analysis of variance was conducted on the log scores. The results are summarized in Table II. All main effects, six two-way interactions, three three-way interactions, and two four-way interactions were significant.

The means for all fixed main effects are given in Table III. The significance of the differences between pairs of means was computed with the Duncan multiple-range test.[3] All the values were significantly different from one another except the three- and four-line conditions in the number-of-lines variable.

Two of the important two-way interactions are shown in Tables IV and V. From Table IV it is clear that for either two, three, or four lines, the multiple-line display is superior in performance to the multiple-graph display. It is interesting to note than an increase in the number of lines had no effect on multiple lines, while it did produce increasingly poorer performance for multiple graphs. Table V indicates that for the point-reading task there is no difference between multiple lines and multiple graphs but that the multiple line is superior for the comparing task.

The only significant subject interaction was subject × display condition. All subjects performed better with the multiple-line display than they did with the multiple-graph display, but for some subjects this superiority was greater than for others.

Of the three significant three-way interactions, the one which is of most interest for comparing multiple-line and multiple-graph displays is the display × confusion × task interaction. This three-way interaction is presented in Table VI. Examination of this table indicates that for the point-reading task, the multiple-line

TABLE III

Geometric Means for Fixed Main Effects

(*a*) *Type of display*

Multiple line	*Multiple graph*
19.18	23.25

(*b*) *Number of lines*

2	*3*	*4*
20.37	21.36	21.64

(*c*) *Confusion among lines*

Low confusion	*High confusion*
20.01	22.28

(*d*) *Task*

Point reading	*Comparing*
18.97	23.51

(*e*) *Line coding*

Black and white	*Color*
21.64	20.61

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.

TABLE IV

Geometric Means for Type of Display by Task

	Task	
Type of display	*Point reading*	*Comparing*
Multiple line	18.95	19.40
Multiple graph	18.97	28.49

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.

TABLE V

Geometric Means for Type of Display by Number of Lines

	Number of lines		
Type of display	2	3	4
Multiple line	19.42	19.30	18.81
Multiple graph	21.36	23.64	24.88

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.

display affects performance adversely with increased confusion, whereas the multiple graph does not. For the comparing task, both types of display conditions affect performance adversely by the addition of confusion.

The two four-way interactions contained all the variables except subjects. These two four-way interactions are shown in Tables VII and VIII, organized so that differences between the multiple-line and multiple-graph presentations are apparent.

It can be seen in Table VII that for the point-reading task, for either high or low confusion or two, three, or four lines, there is essentially no difference between performance using multiple-line or multiple-graph presentation. For the comparing task, multiple lines are significantly better than the multiple graphs for both low and high confusion, with the difference generally becoming greater under the high confusion and under the greater number-of-lines conditions.

Examination of Table VIII reveals some additional information when color and black-and-white line codes are considered. For point reading under low confusion, there are no significant differences between multiple lines and multiple graphs, and color improves performance for both. For high confusion, there are no significant differences between multiple-line and multiple-graph displays; however, color improves the multiple-line but not the multiple-graph presentation.

Under the comparing task, multiple lines are significantly better than multiple graphs for both levels of confusion as well as for both line-coding methods.

The results of the preference ranking are shown in Table IX. For the point-reading task, the subjects generally preferred the multiple-line

TABLE VI

Geometric Means for Display by Confusion by Task

Type of display	*Task*			
	Point reading		*Comparing*	
	Low confusion	*High confusion*	*Low confusion*	*High confusion*
Multiple line	18.08	19.87	18.36	20.50
Multiple graph	18.94	19.00	25.49	31.85

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.

TABLE VII

Geometric Means for Display by Number of Lines by Confusion by Task

	Point reading						*Comparing*					
	Low confusion			*High confusion*			*Low confusion*			*High confusion*		
Type of display	2 *Lines*	3 *Lines*	4 *Lines*	2 *Lines*	3 *Lines*	4 *Lines*	2 *Lines*	3 *Lines*	4 *Lines*	2 *Lines*	3 *Lines*	4 *Lines*
Multiple line	18.14	18.15	17.95	19.61*	18.81	21.28	18.94	19.79	16.52	21.12	20.54	19.86
Multiple graph	17.75	19.29	19.85	18.78*	18.01	20.27	23.83*	27.30	25.42	26.18	32.94	37.46

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.
* This value also not significantly different from 4-line value.

TABLE VIII

Geometric Means for Display by Confusion by Task by Line Code

Type of display	*Point reading*				*Comparing*			
	Low confusion		*High confusion*		*Low confusion*		*High confusion*	
	Black and white	*Color*	*Black and white*	*Color*	*Black and white*	*Color*	*Black and white*	*Color*
Multiple line	18.89	17.30	20.77	19.02	19.38	17.40	20.46	20.53
Multiple graph	19.75	18.16	19.36	18.65	25.39	25.57	31.79	31.91

Notes:
(1) Means reported in seconds.
(2) Values connected by line are not significantly different at $p < 0.05$.

TABLE IX

Preference Ranking for Displays and Line Code by Task

Point reading			*Comparing*		
Display	*Line Code*	*Mean rank*	*Display*	*Line code*	*Mean rank*
Multiple graph	Colored	1.7	Multiple line	Colored	1.2
Multiple line	Colored	2.2	Multiple line	Black and white	1.8
Multiple graph	Black and white	2.7	Multiple graph	Colored	3.3
Multiple line	Black and white	3.4	Multiple graph	Black and white	3.7

1=Most preferred.
4=Least preferred.

and multiple-graph displays in color to their black-and-white counterparts. For the comparing task, the multiple-line presentation in either color or black and white was preferred to the multiple-graph presentation. These findings follow the same general trends as found in the time-score analysis.

DISCUSSION AND CONCLUSIONS

The results of the time-score analysis indicate that for the most accurate prediction of performance, consideration should be given to the two four-way interactions: graph × task × number of lines × confusion, and graph × task × confusion × code. The results for both interactions reveal that the display method has little bearing on the point-reading task. For the comparing task, however, the multiple-line display is always superior. The preference results lend weight to these findings.

Inasmuch as the use of color is always a perplexing problem to the design engineer, perhaps the results for this variable should be examined more closely. In general, it can be concluded that color improves performance for the point-reading task but not for the comparing task. This difference, however, is small in an absolute sense. The greatest effect that color had was in improving performance with the multiple-line display. The finding that color did not have a more profound absolute effect can be partly attributed to the fact that the black-and-white line codes were selected to be an optimally discriminable set. Thus, it appears that color, with its usual concomitant increase in cost and time, is not necessary for good operator performance if a highly discriminable black-and-white line code is utilized.

The specific conclusions from this study are:

(1) If information about the task to be re-

quired of the operator is not available, the multiple-line display is to be recommended over the multiple-graph display.

(2) If the task is known, either the multiple-line or multiple-graph display can be used for the point-reading task. The multiple-line display, however, is recommended for the comparing task. If both tasks are involved, then the multiple-line display is superior for overall performance.

(3) If cost or time to process visual materials are not important factors, color may be used for coding purposes to improve performance slightly.

REFERENCES

1. ANDERSON, C. D., *The Constant-Ratio Rule as a Predictor of Confusion Among Visual Stimuli of Quick Exposure Duration,* AFCRC–TN–58–60, Cambridge Applications Laboratory, Air Force Cambridge Research Center, May, 1959.

2. CONOVER, D. W., and KRAFT, C. L., *The Use of Color in Coding Displays,* WADC Technical Report 55–471, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

3. DUNCAN, D. B., Multiple Range and Multiple F Tests. Reprint Series No. 49; reprinted from *Biometrica,* **11** (1), 1955.

4. HITT, W. D., *et al.,* Development of Design Criteria for Intelligence Display Formats (present edition of this journal).

5. SCHUTZ, H. G., Experiment II: An Evaluation of Formats for Graphic Trend Displays (present edition of this journal).

An Evaluation of Five Different Abstract Coding Methods—Experiment IV[1]

WILLIAM D. HITT, Battelle Memorial Institute, Columbus, Ohio

The purpose of this study was to ascertain the relative effectiveness of selected abstract coding methods, based upon their effects on various operator tasks. Five different coding methods were selected: numeral, letter, geometric shape, color, and configuration. Secondary variables included in the study were target density, number of code levels, and operator tasks. It was found that numeral coding and color coding are the two superior coding methods. If greater emphasis is to be placed on identifying symbols, numeral coding is superior to color coding. No significant differences were found, however, between numeral coding and color coding for the remaining operator tasks: locating, counting, comparing, and verifying.

INTRODUCTION

Abstract coding is generally accepted as a useful technique for representing targets on visual displays. Introducing realism in coding by means of either words or symbols is often not practical or even possible. It would be almost impossible, for instance, to incorporate a high degree of realism in coding the many types of radar sites, missile sites, or aircraft portrayed on visual displays in intelligence centers. For one reason, the differences between items within a single category are frequently too subtle to be portrayed pictorially. Moreover, to spell out all of these items in a readable manner would require considerable space on the display. It therefore seems essential that some form of abstract coding be used.

A number of studies have been conducted to compare code levels within a single code category. For instance, experiments have been conducted to find an optimum set of discriminable colors. Work also has been done on the discriminability of letters and numbers. Moreover, research has been conducted to determine the degree of confusion among various geometric shapes. On the basis of such research, it is now possible to obtain satisfactory discriminability for code levels *within* selected code categories. Very little work has been done, however, to compare the effectiveness of one code category with the effectiveness of a second code category, such as color coding versus geometric-shape coding, or numeral coding versus letter coding.

A review of the literature uncovered only three studies that were particularly pertinent to the comparison of code categories. Eriksen[5] investigated speed in locating objects on a visual display when the various classes of objects on the display differed from one another on only one of the four dimensions of form, hue, size, and brightness. He found that location time for hue differences was significantly shorter than for form differences, and that location times for hue and form were significantly shorter than for either brightness or size.

In a study by Anderson and Fitts,[2] groups of homogeneous color or numeral symbols were exposed tachistoscopically for 0.1 sec. Maximum average information transmitted about color was 10.44 bits per exposure, and maximum average information transmitted about numerals was 14.94 bits.

Alluisi and Muller[1] studied the information-handling performance of subjects with conventional Arabic numerals and six other symbolic visual codes under both motor (key pressing) and verbal-response conditions. The six symbolic codes included a set of straight-line symbolic Arabic numerals, three sets of ordered symbols based upon differences in the visual inclination of a line, a set of colors, and a set of ellipses of differing axis ratios. In a first experiment, subjects responded by pressing one of a number of finger keys for each of the symbols presented. In a second experiment,

[1] This work was sponsored by the U.S. Air Force under Contract AF 30(602)-2078 (ref. RADC–TR–60–201).

subjects responded verbally by calling out the number assigned to each stimulus symbol. The seven codes fell into three clearly different groups under each response condition: the two numeral codes were superior to the three inclination codes, and these were all superior to the two remaining codes of color and ellipse-axis ratio.

Thus, even though only a limited number of studies have been conducted to evaluate the relative effectiveness of display categories, the results indicate at least one apparent inconsistency. Eriksen found hue to be superior to form in coding objects on visual displays. Anderson and Fitts, as well as Alluisi and Muller, on the other hand, found numerals to be superior to color.

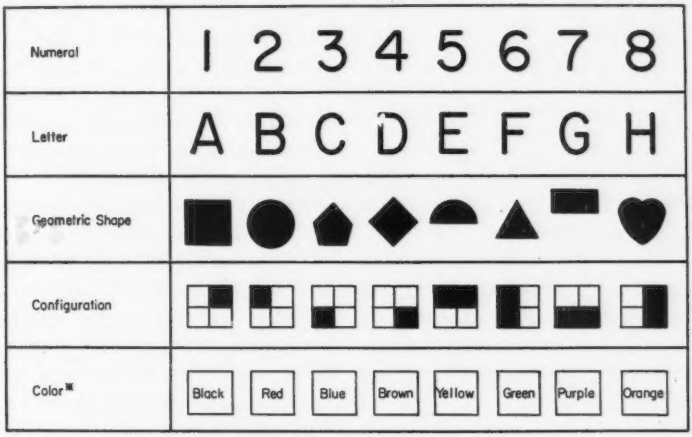
One possible explanation for these discrepant results is that the task required of the subjects was not the same in all three studies. The subjects in Eriksen's study were required to locate objects, whereas the subjects in the other two studies were required to identify objects. On the basis of the difference in task required of the subject, it is indeed plausible to assume a significant interaction between visual display conditions and operator tasks. A review of the literature on visual displays did not reveal any previous attempt to test this conjecture.

The objective of this study, then, was to ascertain the relative effectiveness of selected abstract coding methods, based upon their effects on various operator tasks.

METHOD

Design of Experiment

Three major steps were involved in the design of the experiment. First, it was necessary to decide which variables to include in the experiment. Next, it was necessary to specify the levels or steps over which to manipulate the selected variables. The final task was to organize these variables in a suitable manner to produce a well-controlled experimental design.



* No attempt was made to hold saturation and brightness constant

Fig. 1. The five code categories.

Five different coding methods were selected: numeral, letter, geometric shape, color, and configuration. The first four coding methods are rather obvious techniques to consider, and the fifth is a modification of the inclination codes used by Alluisi and Muller.[1] On the basis of experimental evidence which indicates that no more than approximately nine different hues can be discriminated on an absolute basis,[4] the upper limit for number of code levels was

chosen as eight. The eight levels for each of the five code categories are presented in Fig. 1.

To achieve a high degree of generality in the evaluation of these five coding methods, the number of code levels and the target density were included in the study as secondary independent variables. The number of code levels was varied over two, four, and eight levels. Target density was varied over 40, 80, and 120 symbols per display.

Since the objective of the study was to determine the effectiveness of selected abstract coding methods in relation to operator tasks, operator task was also included as an independent variable. According to a somewhat general analysis of the system under study, five different operator tasks appeared to be basic to visual display reading. These tasks were: identifying, locating, counting, comparing, and verifying.

It was decided to select five males from Battelle's professional research staff to serve as subjects in the experiment.

In order to test for important interactions between variables, the experiment was designed as a complete factorial for the variables described above. That is, with subjects being "random" and all other variables being "fixed", the experimental design was a $5 \times 3 \times 3 \times 5 \times 5$ factorial (5 coding methods, 3 code levels, 3 densities, 5 tasks, and 5 subjects).

Inasmuch as both speed and accuracy of performance are important for the system under study, number of correct responses per experimental condition and speed of response per experimental condition were selected as the measures of operator performance.

Information believed to be representative of the inputs into an intelligence center was developed in the form of eight items under each of five target categories: Industry, Personnel Buildings, Radar Sites, Aircraft, and Defense Units. These target categories, which constitute the information to be represented by the abstract codes, are presented in Table I. All five target categories were included under each of the 225 experimental conditions.

To complete the experimental design, several additional control variables were taken into consideration. These control variables were incorporated in either a random or a partially counterbalanced manner. The counterbalancing technique was used so that the same code or the same test category would not be presented to the same subject on two successive occasions, which might have resulted in unwanted transfer effects.

To control for warm-up and fatigue effects within a single experimental session, the order of presentation of experimental conditions within a session was completely randomized. To control for day-to-day learning effects, as well as for other causes of intrasubject variability, the order of presentation for experimental conditions over sessions was partially counterbalanced.

Presentation of Materials

To eliminate the possibility of favoring one code category over another, available human-engineering design principles were applied to the design of all displays. The selection of

TABLE I

Target Categories

Item	*Industry*	*Personnel buildings*	*Radar sites*	*Aircraft*	*Defense units*
1	Aluminium plant	Command post	Early warning	Interceptor	Army unit
2	Petroleum refinery	Barracks	Antiaircraft	Short-range bomber	Navy unit
3	Gas-producing plant	Headquarters	Aircraft control and warning	Long-range bomber	Air Force unit
4	Lead refinery	Prison	Gap filler	Tanker	Coast Guard unit
5	Steel mill	Hospital	Texas tower	Transport	Marine unit
6	Aircraft assembly plant	Theater	Airborne search	Trainer	Civil Defense unit
7	Explosives plant	Mess hall	Picket line	Civil	National Guard unit
8	Cement plant	School	Civil surveillance	Helicopter	Military Reserve unit

colors, configurations, and geometric shapes, for instance, was based upon the results of previous research; the size of symbols—½ in. for longest dimension—was based upon specific human-engineering recommendations.[3]

The displays were 20 × 13-in. rectangles mounted on 30 × 22-in. cardboard posters with eight columns and five rows. Coördinates for the 40 cells were indicated by 1-in. red numerals along the top and bottom borders of the

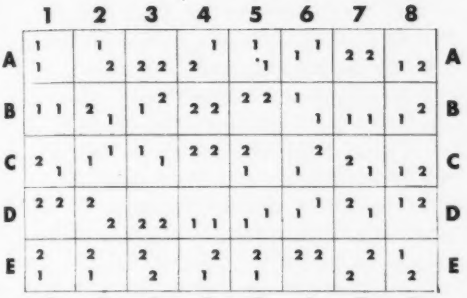


Fig. 2. Numeral coding: two levels, intermediate density.

display, and 1-in. red letters along the sides. Targets were laid out in a rectangular grid system.

For each of the five code categories nine display posters were prepared (i.e. for three target densities × three code levels), resulting in a total of 45 display posters. All cells within a display contained one, two, or three symbols,

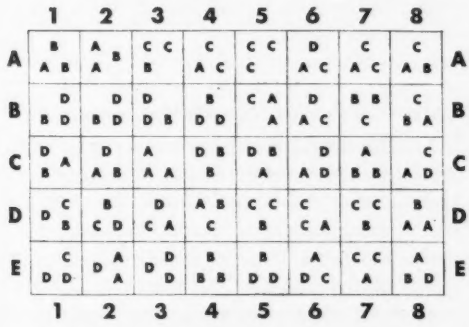


Fig. 3. Letter coding: four levels, high density.

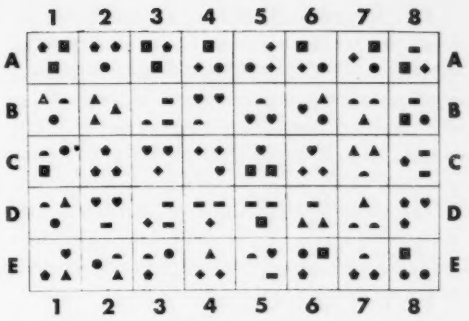


Fig. 4. Shape coding: eight levels, high density.

depending upon the density level being represented. The location of a symbol within a cell was determined in a random manner. The types of targets within a code category were represented an equal number of times on all displays. All nine displays of one code category corresponded exactly to the nine displays for any of the four other code categories, except,

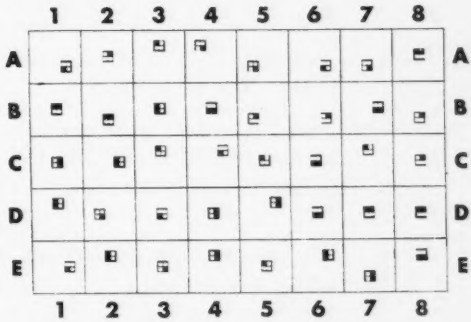


Fig. 5. Configuration coding: eight levels, low density.

of course, for the type of targets. Examples of the 45 display posters are presented in Fig. 2 through 5.

A set of 2 × 2-in. training cards was prepared for each of the five code categories. Except for being larger in size, the symbols on these cards were exact replicas of the target symbols on the posters. The cards were constructed for the purpose of teaching subjects how to encode and decode the target categories.

Pencil-and-paper tests were constructed to measure the five operator tasks. A subset of questions for a single display poster consisted of 25 questions—five questions for each of the five tasks. Thus, for a set of nine display posters, a basic set of 225 questions was constructed. This basic set of 225 questions was written five different times to constitute five equivalent forms of the same test—one form for each of five target test categories. Examples of the questions are listed in Table II.

TABLE II

Examples of Questions to Measure Five Operator Tasks

1. *Identify:* What type of aircraft is in 4-C?
   (a) Interceptor.....
   (b) Short-range bomber.....
   (c) Long-range bomber.....
   (d) Tanker .....
2. *Locate:* Which column contains one cell with only early warning radar sites?.....
3. *Count:* How many aluminium plants are in Row C? .....
4. *Compare:* Does 2-D contain the same composition of defense units as 8-C?
   Yes .....
   No .....
5. *Verify:* The personnel buildings in 4-D are command posts.
   True .....
   False .....

Experimental Procedure

The subjects were tested for visual acuity and color vision prior to the first experimental trial. Detailed instructions were given to the subjects, at which time the experimenter emphasized the importance of both speed and accuracy of performance.

The first part of each experimental session involved teaching the subject the associations for the code category/target category combination which was specified for that particular day. The experimenter stated the associations for the entire sequence twice in succession, showing the subject the training card as he stated the corresponding item in the target category; for example, "Black represents an interceptor, red represents a short-range bomber, blue represents a long-range bomber, etc." The subject then practiced identifying the codes until he achieved a 100 per cent criterion twice in succession.

Next, the subject was instructed to read the five test items on the first page of a test booklet. If he had any questions, they were answered by the experimenter. When the subject signaled he was ready to commence, the experimenter raised a blind that covered the display poster and, simultaneously, started the timer. Immediately after the subject completed the fifth question, he again signaled to the experimenter, who stopped the timer, lowered the blind, and then recorded the time. This procedure was followed throughout an experimental session.

During each experimental session, a subject completed 15 experimental conditions: five tasks × three target densities, with code level, code category, and target test category held constant. An experimental session lasted between 30 and 45 min. At the rate of 15 conditions per session, 15 experimental sessions (one per day for 15 successive work days) were required for a subject to complete the 225 experimental conditions. At the completion of the experiment, all five subjects had completed the necessary trials for all 225 experimental conditions.

Cross-Validation Experiment

In order to obtain an estimate of the amount of error in the experiment and to better characterize the population of subjects from which the original sample of five was drawn, the entire experiment was repeated on a second group of five subjects. This cross-validation experiment was conducted in the same manner as the original experiment, except for two

changes: (1) a different group of five subjects (who were also professional research scientists representing various scientific disciplines) were selected, and (2) slides were used rather than posters for presentation of stimulus materials. This latter change was made because of the noticeable wear on the posters after the completion of the first experiment.

RESULTS

To achieve an approximately normal distribution of scores, as well as to make use of both performance measures, accuracy scores and speed scores were combined in the form of "number of correct responses per minute". The product-moment correlation between the accuracy scores and speed scores for the first replication was −0.82, but, nonetheless, when plotted alone, each measure formed a highly skewed distribution. Combining the scores, however, resulted in both normality of distribution for all scores combined and homogeneity of variance for the experimental conditions.

An estimate of intersubject reliability for the first experiment was obtained by computing the average correlation coefficient between subjects over the 225 experimental conditions. This correlation was found to be 0.78, which indicates a rather high degree of consistency in performance among the five subjects.

INTERCORRELATION AND FACTOR ANALYSIS FOR OPERATOR TASKS

It was next desired to determine the degree of relation between the five operator tasks. By treating the mean scores for subjects as individual scores, the five tasks were intercorrelated over the remaining 45 experimental conditions. These intercorrelations are reported in the upper right portion of Table III.

In order to determine the extent to which these five tasks could be reduced to a smaller number of basic dimensions, the matrix of intercorrelations for the first experiment was factor analyzed by means of the centroid method. The two factors presented in Table III (b) were obtained from this analysis. By use of simple structure as the criterion for rotation, these two orthogonal factors were rotated to produce the transformed factors in Table III (d). Fig. 6 represents a graphic plot of the two rotated factors. It is clearly seen from this figure that locating and identifying (or *search* and *recognition*) represent two independent operator tasks. On the other hand, the verifying, counting, and comparing tasks are loaded on both of these factors.

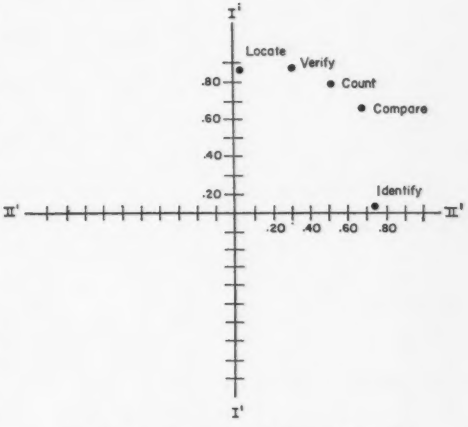


Fig. 6. A graphic plot of rotated factors.

Analysis of Variance

As mentioned previously, the experimental design was a $5 \times 5 \times 3 \times 3 \times 5$ factorial, with subjects considered "random" and the other variables considered "fixed". The complete analysis of variance, which tested for all main effects and for all interactions except the five-way interaction, is presented in Table IV.

The relative effectiveness of the five code categories, based upon their effects on operator tasks, is shown in Table V. The code categories are rank ordered on each of the tasks according to the mean correct-response-per-minute scores. Code categories connected by the horizontal lines are not significantly different at $p < 0.05$.

TABLE III

Intercorrelation Matrix and Factor Matrix for Tasks (for First Replication)

(a) *Intercorrelations and residuals*

	Identify	Locate	Count	Compare	Verify
Identify		0.02	0.47	0.52	0.28
Locate	−0.02		0.72	0.59	0.74
Count	0.03	0.03		0.89	0.83
Compare	−0.01	0.01	−0.01		0.81
Verify	0.02	−0.02	−0.03	0.02	

Correlations: upper right; residuals: lower left.

(*b*) *Unrotated factors*

	I	II
Identify	0.46	0.56
Locate	0.71	−0.50
Count	0.96	0.00
Compare	0.94	0.18
Verify	0.89	−0.26

(*c*) *Transformation matrix* (37 *degrees clockwise*)

$$X \begin{vmatrix} 0.80 & 0.60 \\ -0.60 & 0.80 \end{vmatrix} =$$

(*d*) *Rotated factors*

	I′	II′	h^2
Identify	0.03	0.73	0.53
Locate	0.87	0.02	0.76
Count	0.77	0.57	0.92
Compare	0.64	0.70	0.91
Verify	0.87	0.33	0.87

It can be seen from Table V that, in general, color coding and numeral coding are the two superior coding methods. For the identifying task, however, numeral coding is ranked first and color coding is ranked fourth. On the other hand, color coding is not significantly different from numeral coding on the remaining tasks.

Tables VI and VII indicate, respectively, the effect of number of code levels and the effect of target density on operator performance. As might be expected, increases both in number of code levels and in target density degrade operator performance. It is interesting to note that, in both cases, the decrease in performance from the first level to the second level is relatively large, and then from the second level to the third level there appears to be a leveling-off effect.

Cross-Validation Experiment

A product-moment correlation between mean subject performance for the first experiment and mean subject performance for the second experiment, based upon an N of 225 experimental conditions, was found to be 0.97. With such a small amount of error in the experiment, it is apparent that, for all practical purposes, any interpretation made of the results

TABLE IV

Summary of Analysis of Variance

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
Subjects (S)	4	256.7973	Residual	153.77*
Codes (C)	4	390.4855	SC	17.78*
Levels (L)	2	404.4655	SL	99.87*
Density (D)	2	1456.7000	SD	156.63*
Tasks (T)	4	1649.4250	ST	182.46*
SC	16	21.9554	Residual	13.15*
SL	8	4.0504	Residual	2.42*
SD	8	9.2954	Residual	5.57*
ST	16	9.0401	Residual	5.41*
CL	8	14.6700	SCL	—
CD	8	.8200	SCD	—
CT	16	24.9318	SCT	7.99*
LD	4	10.8688	SLD	4.49*
LT	8	359.0986	SLT	73.89*
DT	8	78.6800	SDT	28.10*
SCL	32	31.1250	Residual	18.63*
SCD	32	4.2586	Residual	2.55*
SCT	64	3.1227	Residual	1.87*
SLD	16	2.4152	Residual	1.45*
SLT	32	4.8643	Residual	2.91*
SDT	32	2.8027	Residual	1.68*
CLD	16	2.4400	SCLD	—
CLT	32	2.8900	SCLT	—
CDT	32	2.1627	SCDT	1.14
LDT	16	71.4916	SLDT	20.66*
SCLD	64	3.5725	Residual	2.14*
SCLT	128	3.1179	Residual	1.87*
SCDT	128	1.8964	Residual	1.14
SLDT	64	3.4657	Residual	2.07*
CLDT	64	2.0920	Residual	1.28
Residual	256	1.6725		

* $p < 0.05$.

from the first experiment also would apply to the results from the second experiment. For this reason, the results from the cross-validation experiment were not analyzed in any great detail.

Because of the special significance of the factor analysis of operator tasks, however, it was decided to factor analyze the tasks for the cross-validation experiment in the same manner that the factor analysis was conducted for the first experiment. A comparison of Table III with Table VIII reveals the high degree of similarity between both the intercorrelation matrices and the factor matrices for the two replications, as would be expected from the above-mentioned correlation of 0.97.

TABLE V

The Effect of Coding Methods on Operator Tasks

Tasks	*Rank order of code categories*				
	First	*Second*	*Third*	*Fourth*	*Fifth*
Identify	Numeral	Letter	Shape	Color	Configuration
	13.64	13.02	12.53	12.34	11.77
Locate	Color	Numeral	Letter	Shape	Configuration
	8.46	7.42	7.25	6.94	4.03
Count	Numeral	Color	Shape	Letter	Configuration
	12.60	12.22	11.49	11.11	7.07
Compare	Numeral	Color	Shape	Letter	Configuration
	6.85	6.72	6.56	6.33	4.76
Verify	Numeral	Color	Shape	Letter	Configuration
	10.01	9.95	9.50	9.05	6.60

Note:
(1) Scores reported in terms of mean correct response per minute.
(2) Code categories connected by line are not significantly different at $p < 0.05$.

TABLE VI

Effect of Number of Code Levels on Operator Performance

Number of code levels	*Mean number of correct responses per minute*
2	10.31
4	8.76
8	8.34

TABLE VII

Effect of Target Density on Operator Performance

Target density (Number of symbols per display)	*Mean number of correct responses per minute*
40	11.36
80	8.45
120	7.60

DISCUSSION AND CONCLUSIONS

One of the most important findings of the present study was the identification of recognition and searching as two independent operator tasks. There are several reasons why this finding is particularly important. First, it suggests a possible explanation for the seemingly discrepant results of previous research. Second, it helps to explain the obtained interaction between code categories and operator tasks. Third, it strongly implies that recommendations given to the design engineer concerning the design of visual displays must be based upon knowledge of the tasks involved in the system. Then, finally, the identification of basic task factors should help establish a more rigorous framework for future research on visual displays.

It was pointed out above that, in general, color coding and numeral coding were superior to the other coding methods. This finding, of course, was based upon an upper limit of only eight code levels. According to the results of previous research on absolute identification of hue, it would seem highly likely that numeral coding—or even one of the other coding methods—would be superior to color coding if more than nine or ten code levels were desired.

The correlation of 0.97 between replications permits a rather high degree of assurance in

TABLE VIII

Intercorrelation Matrix and Factor Matrix for Tasks (for Second Replication)

(*a*) *Intercorrelations and residuals*

	Identify	Locate	Count	Compare	Verify
Identify		0.00	0.53	0.56	0.23
Locate	−0.03		0.65	0.45	0.67
Count	0.03	0.06		0.80	0.69
Compare	0.00	−0.05	−0.04		0.73
Verify	0.02	−0.04	−0.03	0.07	

Correlations: upper right; residuals: lower left.

(*b*) *Unrotated factors*

	I	II
Identify	0.50	0.60
Locate	0.68	−0.51
Count	0.92	0.06
Compare	0.89	0.20
Verify	0.81	−0.32

(*c*) *Transformation matrix* (38 *degrees clockwise*)

$$X \begin{vmatrix} 0.79 & 0.62 \\ -0.62 & 0.79 \end{vmatrix} =$$

(*d*) *Rotated factors*

	I′	II′	h^2
Identify	0.03	0.78	0.61
Locate	0.85	0.02	0.72
Count	0.69	0.62	0.85
Compare	0.58	0.70	0.83
Verify	0.83	0.25	0.76

making recommendations from the results of the study. It is apparent that the results of the second replication could have been predicted almost perfectly from the results of the first replication. Inasmuch as different subjects were used in the second replication and posters were replaced by slides (which produced certain noticeable changes in the displays), a certain amount of generalizing is warranted.

The specific conclusions from the study are:

(1) Searching and recognition are two independent task factors. This finding should be kept in mind when designing future experiments to evaluate visual displays, as well as when making recommendations concerning the design of visual displays.

(2) Within the limits of the present study, color coding and numeral coding are the two superior coding methods. If greater emphasis is to be placed on recognition of symbols, however, numeral coding is superior to color coding. For the remaining tasks, there is no significant difference between color coding and numeral coding.

REFERENCES

1. ALLUISI, E. A., and MULLER, P. F., Jr., "Verbal and Motor Responses to Seven Symbolic Visual Codes: A Study of S–R Compatibility". *J. Exp. Psychol.* **55**, 247–254, 1958.

2. ANDERSON, N. S., and FITTS, P. M., "Amount of Information Gained During Brief Exposures of Numerals and Colors". *J. Exp. Psychol.* **56**, 362–369, 1958.

3. BAKER, C. A., and GRETHER, W. F., *Visual Presentation of Information,* WADC Technical Report 54–160, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

4. CONOVER, D. W., and KRAFT, C. L., *The Use of Color in Coding Displays,* WADC Technical Report 55–471, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

5. ERIKSEN, C. W., "Location of Objects in a Visual Display as a Function of the Number of Dimensions on which the Objects Differ. *J. Exp. Psychol.,* **44**, 56–60, 1952.

An Evaluation of the Effect of Selected Combinations of Target and Background Coding on Map-Reading Performance—Experiment V[1]

CHARLOTTE A. CHRISTNER, HORACE W. RAY, Battelle Memorial Institute, Columbus, Ohio.

The objective of this study was to determine the relative effectiveness of selected target-background coding combinations. Three target codes were: color, number, and enclosed shape. Five types of background were: all white, solid gray, five shades of gray, five pastel hues, and five different patterns. These target-background coding combinations were evaluated under eight different complexity conditions and for five different operator tasks. The major findings in this study were: (1) no significant differences were found in background coding, (2) numeral coding is superior for the "identifying" task, and (3) color coding is superior for the "locating" and "counting" tasks.

INTRODUCTION

The coding of a cartographic display is difficult for a number of reasons:

(1) The amount of information displayed is necessarily relatively high because two sets of data (subject data and geographic data) are presented simultaneously.
(2) Data are presented concurrently in an area of limited size, with severe constraints on the designer's freedom to distribute the data over that space.
(3) The number of kinds and levels of information to be displayed is sufficiently high to tax the ingenuity of the designer in preparing a matching number of perceptual dimensions and steps.
(4) The trade-off values are hardly known for
   (a) Increasing both the number of coding dimensions and steps simultaneously, versus
   (b) Combining or summarizing the basic information.
(5) The arrangement of the data in a figure-ground relationship raises possibilities of significant interactions which have been found in the studies of perception.

The human-engineering evaluation of cartographic displays is also a problem. Evaluations customarily are based on such tasks as single-target recognition or visual search, with extremely high speed and accuracy requirements. While such tasks are undoubtedly elements of a more comprehensive map-reading or map-interpretation task, they are surely not the only elements. Furthermore, the usual criteria may be unrealistically high considering the over-learning and the availability of such simple aids as a key or legend which may be present in the real situation.

Unfortunately, neither display "complexity" nor map-reading performance are understood well enough to permit specification of the coding dimensions and steps to be used for a cartographic display. Two general stipulations seem to be in order:

(1) Differences between figure and ground, among figures, and among background areas must somehow all be maintained at a high level at the same time.
(2) There is probably some point in increasing display complexity beyond which additional coding dimensions are a hindrance and the best coding dimensions are the simplest ones.

The aim of this study was to evaluate the relative effectiveness of the combinations of certain selected coding dimensions for targets

[1] This work was sponsored by the U.S. Air Force under Contract AF 30 (602)–2078 (ref. RADC–TR–60–201).

and for backgrounds. Multi-dimensional codes were not considered in this study.

METHOD

Fifteen simple coding combinations chosen for apparent human-engineering qualifications, were presented with different display complexity levels and tasks to provide greater confidence in the generalizability of the results. Specifically, three variables considered representative of complexity were presented at two levels each, and the experimental task included five subtasks.

Because of the importance usually attached to color for attention-getting, and because of the technical difficulties in high-speed production of colored displays, a major objective of the study was to determine the merits of color coding relative to black-and-white coding techniques. Consequently, color coding was used both for targets and for backgrounds.

TABLE I

Munsell Notations for Colors Used in Target Coding*

Book notation	*Munsell production number*
1R†	999
9R	892
1Y†	946
7GY	960
9G†	1099
5B	1087
1P†	1135
3RP	1003

* All colors in this set are from Munsell series of 50 maximum chroma.
† Colors used for sets of four levels of target coding.

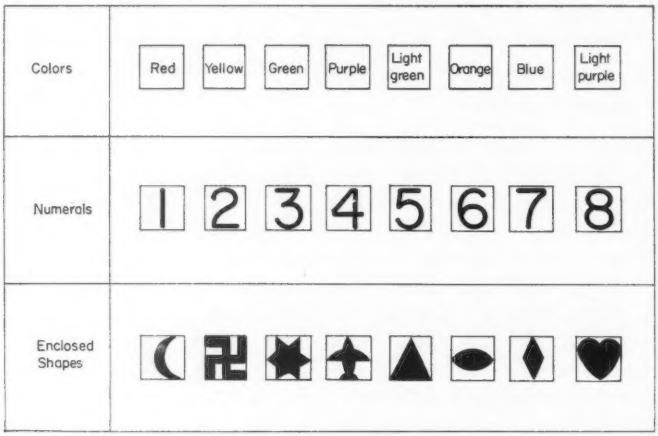


Fig. 1. The three target codes.

Target Coding

Colors used for target coding were those recommended by Conover and Kraft,[4] for eight- and four-step codes. (Munsell notations are indicated in Table I.) The targets were $\frac{1}{2} \times \frac{1}{2}$-in. squares with a hairline black border. The targets have approximately the same saturation and brightness, and therefore the major distinguishing cues are differences in hue.

Numbers were selected in preference to plain shapes as a technique for target coding, on the basis of their previously demonstrated effectiveness (Hitt[7]).

The design of enclosed shape targets was based on eight of the shapes recommended by Sleight.[1] The solid black shapes were made into square targets by enclosing each in a $\frac{1}{2} \times \frac{1}{2}$-in. white area with a 1-mm black border. Shape sizes were adjusted so that the longest dimension of each shape just fitted within the

borders of the square. Selections were made from Sleight's set in order of best discriminability, except where it seemed that discriminability would be greatly reduced through interaction of shape and border. One shape, the heart, was modified slightly to increase the white area of the square. It was hoped that the enclosure of the shape in a square would enhance contrast effects and provide a helpful redundancy of information (see Fitts),[6] without increasing target area to the extent that shape coding alone, but in larger size, might just as well have been used (see Bridgman and Wade).[2] The three target codes are presented in Fig. 1.

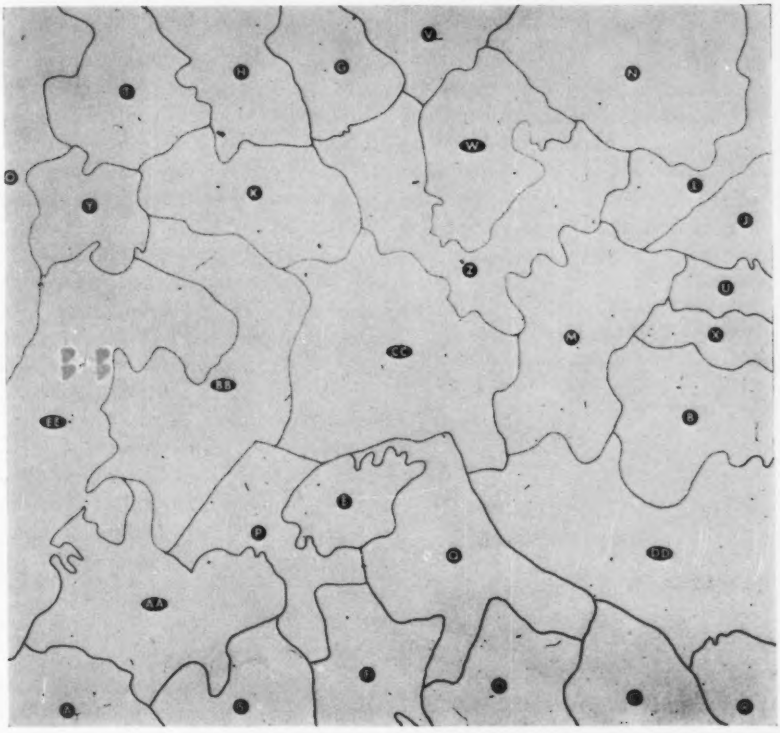


Fig. 2. Basic map areas without targets.

Background Coding

One basic map was used for all five coding conditions. The map was selected from an atlas on the basis of two criteria: (1) it should be unfamiliar to all subjects, and (2) it should contain areas whose shape represented an average degree of irregularity, according to the judgment of the experimenters. The map selected was one of West Central Asia containing 30 areas depicting small sections of Afghanistan, China, Iran, Iraq, Turkey, and the U.S.S.R.

Because the actual area represented was of no significance to the study, no attempt was made to present accuracy of detail. The various provinces were treated as individual units without regard to the country to which they belonged; water was not differentiated from land; and some simplification of borders

resulted from tracing the map. Only sector outlines were traced: cities, rivers, and other map detail were entirely omitted. Map areas were designated by single letters of the alphabet and double letters AA through EE ("I" was omitted because of possible confusion with the number "1"). The area "names" were labelled in $\frac{3}{8}$-in. white letters enclosed in solid black circles or ovals, located approximately in the center of each area. The map is shown in Fig. 2.

The five colors selected for background coding were expected to provide maximum brightness contrast with the targets, while still maintaining high discriminability of map sectors. Colors were assigned to the background areas randomly, with the restriction that no adjacent areas should be the same color. Munsell notations for these colors are presented in Table II.

TABLE II

Munsell Notations for Colors Used in Background Coding

Value/Chroma	*Hue*
8/4	2.5 (red)
8/4	2.5 (yellow)
8/4	2.5 (green)
8/4	2.5 (blue)
8/4	2.5 (purple)

Various patterns such as crosshatch, dots, and diagonal lines are common background codes. These not only make the areas of a map background discriminable, but theoretically may also improve target visibility. The patterns were selected by the experimenters' judgments on discriminability. They were designed so that they were nearly equal in brightness when viewed at a distance at which the patterns were no longer discriminable.

Grays of varying levels were used for brightness coding of the background. In Munsell designation, these were N 9.5/, N 8/, N 7/, N 6/, and N 5/. With the exception of the white (N 9.5/), all are separated by equal intervals of two steps. The white was separated from the nearest gray by three steps to provide a means of showing basic differences in map areas, e.g. water as distinct from land.

Even on maps which do not show areas in different colors, patterns, etc., area coding by shape, location, and usually name still exits. The map usually remains the same while the data symbols superimposed on it vary. Consequently, the most important function of coding the different map sectors might be to show their extent, especially where the borders are irregular or where they are partially obscured by a heavy posting of data. Considering these factors, one questions whether differential coding of map sectors might detract more from efficient performance by reason of additional "clutter" on the display, than it would add to efficiency by increasing the distinctiveness of areas.

Accordingly, two conditions were used in which map areas were distinguished only by shape, location, and name. In one case, an all-gray map was used; in the other, all white. The latter might have sufficed to test for "plain" maps, but observations by the experimenters indicated that a fairer test of plain backgrounds might be furnished by a gray background. Such a background appeared to give a better contrast with two of the target codes: enclosed shape and color. In both cases, the appearance of white or color against the gray background seemed almost luminous.

Complexity

In recognition of the multidimensional nature of perceptual complexity, a number of possible variables were listed, which seemed most likely to interact with target-background coding. The variables were as follows:

Group 1—Target variables

(1) Total number of letters on the map.
(2) Absolute number of targets.
(3) Number of coding dimensions.
(4) Number of coding levels for each dimension.
(5) Total area of targets per total area of display.

(6) Clustering of targets (targets packed closely together versus targets widely separated).

Group 2—Variables relating to linear features

(1) Total length of linear features such as borders, coasts, rivers.

(2) Extent to which linear features intertwine and criss-cross.

(3) Number of coding dimensions.

(4) Number of coding levels for each dimension.

Group 3—Area variables

(1) (Perimeter)²/Area (an index of form complexity).

(2) Number of areas.

(3) Number of coding dimensions.

(4) Number of coding levels for each dimension.

It is evident that this list is not complete, and that much work needs to be done in conceptualizing and defining cartographic density-complexity variables. The aim of the present study, however, is merely to investigate the effects of selected target and map coding combinations on the performance of an assortment of map-reading tasks. Therefore, only three of these variables were used, at two levels each. These were:

(1) Number of targets (60 and 120).

(2) Number of coding levels (4 and 8).

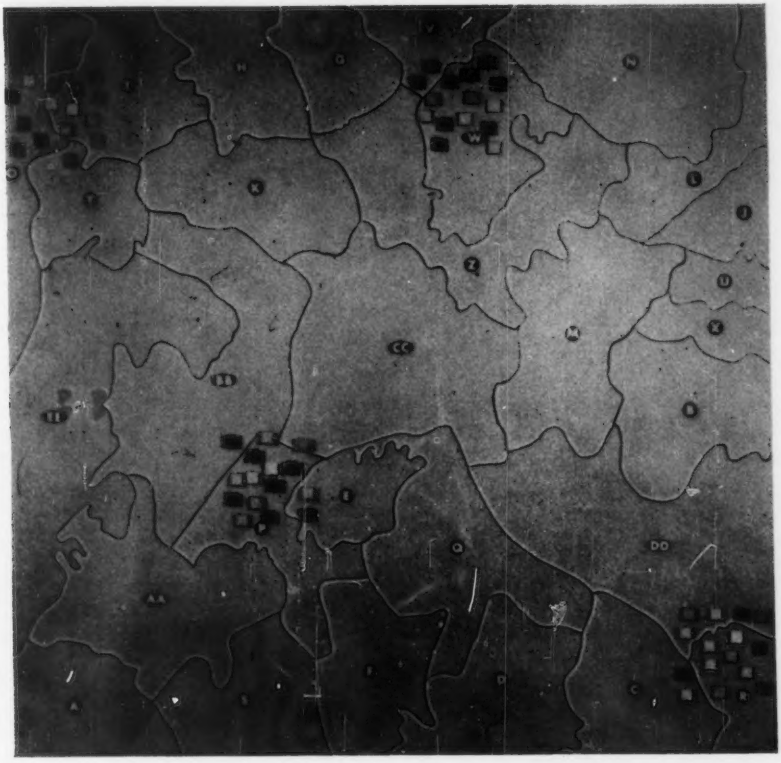


Fig. 3. Solid-gray background: high clustering, four coding levels, 60 targets, color coding.

(3) Clustering of targets (high versus low).

Number of targets. This was simply the grand total of all targets appearing at one time on a 30 × 30-in. display panel. The 60- and 120-target levels were chosen on the basis of previous research[3] as being within a range in which significant effects on performance might be expected.

Number of coding levels. For each of the three coding dimensions for targets, the number of steps used was either four or eight. The eight-level approaches the practical maximum reported for absolute discriminability of hues. For any given display condition, only one target coding dimension was used, but within that dimension, four or eight different kinds of targets appeared.

Clustering of targets. When one looks at a map, it is common to see that some of the targets are packed very closely together to form a target complex or constellation (e.g. a number of NIKE sites placed around a city). At the other extreme, targets are often spread out over the entire map, as might be the case with Air Force bases. These extremes occur, even though the total number of targets on the map remains constant.

In order to vary experimentally the degree to which targets were clustered together, the total number of targets for each condition was

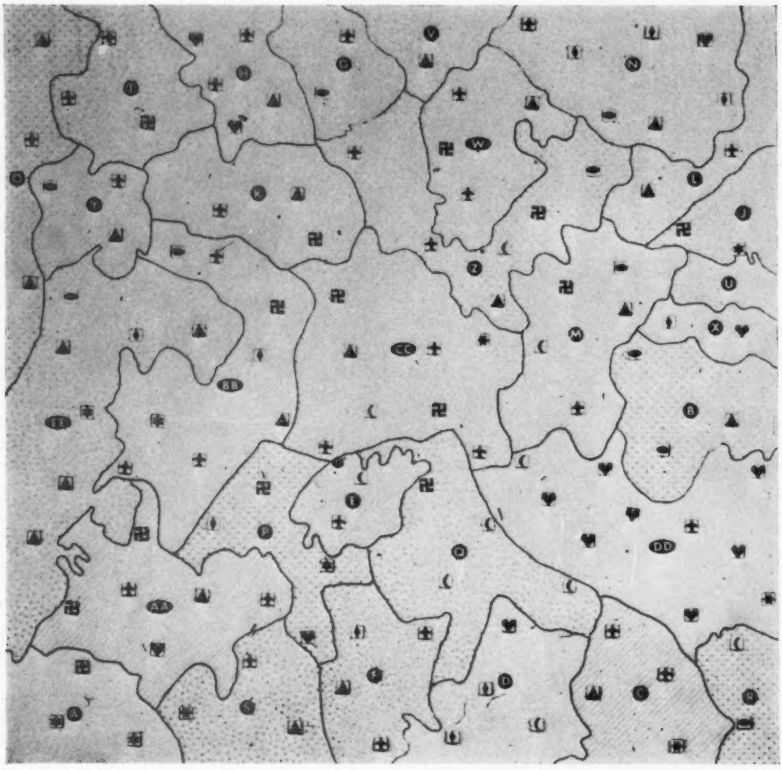


Fig. 4. Patterned background: low clustering, eight coding levels, 120 targets, enclosed-shape coding.

divided into four equal-size groups, one group assigned to each quadrant of the display panel. For each quadrant, targets were placed within a square "sub-area" within the quadrant of a size which clustered targets together closely or clustered them together loosely.

For the high-clustering condition, the square sub-area in which the targets were placed was of a size such that a quarter of this sub-area was covered with targets. For the low-clustering condition, the sub-area in which the targets were placed was of a size such that only a twenty-fifth of this area was covered. The selection of these clustering levels was quite arbitrary, but appeared to the experimenters to result in very crowded and quite uncrowded conditions. Examples of target coding and complexity variables can be seen in Figs. 3 and 4.

Experimental Task

Each of five subjects was given a series of questions designed to sample several subtasks considered basic to map reading. The operations to be performed included locating, identifying, and counting targets, as well as comparing and verifying target data in different areas.

Experimental Design

There were 15 coding combinations representing three target codes and five background codes. The evaluation of these coding combinations was of primary interest. Another set of independent variables was the three complexity variables, represented at two levels each. In addition, there were five operator tasks and five subjects. These variables were varied factorially, yielding a $15 \times 2 \times 2 \times 2 \times 5 \times 5$ design without replication.

Table III presents all combinations of target-background coding and the complexity variables. These combinations represent the 120 experimental conditions associated with changes in the visual display. For each cell, time scores, accuracy scores, and number of responses were recorded for each of the five subjects, the same five subjects performing

TABLE III

The Fifteen Coding Combinations and Eight Complexity Conditions

Number of levels:	4	4	4	4	8	8	8	8
Clustering:	1/25	1/25	1/4	1/4	1/25	1/25	1/4	1/4
Number of targets:	60	120	60	120	60	120	60	120
Background: Target:								
Color Color								
Color Number								
Color Shape								
All Gray Color								
All Gray Number								
All Gray Shape								
White Color								
White Number								
White Shape								
Patterned Color								
Patterned Number								
Patterned Shape								
Diff. Gray Color								
Diff. Gray Number								
Diff. Gray Shape								

under each display condition. The experimental conditions in the table were repeated five times, once for each task.

In addition to the independent variables already discussed, certain control variables were included. First, there were five different categories of information (defense units, army personnel buildings, aircraft, industry, and radar sites), all five of the categories occurring in each cell of Table III. These same five categories are described in detail elsewhere (Hitt[7]). Different categories were used in order to nullify any pre-experimental association between a particular target coding method and a particular category of information.

To prevent the subjects from learning the scheme of arrangement of the overlays and to obviate the effect of unintentional differences in the make-up of overlays for different conditions, two additional control variables were incorporated. First, the location of target clusters in each quadrant of the display was determined by randomization procedures, in order that differences between display conditions could not be attributed to a particular location of targets on the display. Second, the number of each kind of target in each cluster was randomly determined.

Equipment

The five backgrounds were drawn on 30 × 30-in. cardboard. In order to represent all combinations of target coding, background coding, and the complexity variables, it was found expedient to place acetate overlays over the map backgrounds. Targets were made with Scotch-tape backing, and these targets were pasted on to the acetate. A total of 240 acetate overlays were needed in order to represent all of the combinations of the display variables.

Question Booklets

The basic types of questions asked for each task can best be understood by referring to the questions themselves, examples of which are presented in Table IV. The Identify questions were multiple choice, with four alternatives. For Locate, all 30 area designators were listed under each question, and the subject was required to circle all area designators which applied to the question.

Each phrase, number, or area designation which is *italicized* in Table IV indicates that these parts of the questions changed, depending upon the make-up and placement of target clusters, the category of information, and the display condition. All changes were made randomly, so as not to bias effects of the independent variables.

RESULTS

Performance Measures

For the entire experiment, a set of 3000 time scores, 3000 error scores, and 3000 number-of-response scores were obtained. These scores represented 15 coding treatments $\times(2\times2\times2)$ complexity levels $\times 5$ tasks $\times 5$ subjects. Although each task was always represented by three questions, the questions were timed as a unit, and the error scores pooled for each presentation of the task.

In addition to the performance scores, there were five rating-scale scores per coding condition, which indicated the subjects' preferences for the various coding conditions.

An examination of the raw scores revealed that errors were negligible for all tasks except the Count task. Therefore, the primary analysis was conducted on time scores and number of responses. A table of the absolute errors for the Count task was later reviewed. A Friedman chi square computed for the table as a whole was significant, but no further statistical analysis was undertaken, since these data clearly would not produce results contradictory to the findings from the analysis of time scores and number of responses.

A rate-of-responding measure was obtained by dividing the number of responses by the time taken (in minutes). The logarithms of the rate measures were then computed, and used in the statistical analysis. Logarithmic scores were used in order that average rate of respond-

TABLE IV

Questions for the Five Tasks

Identify

1. What type(s) of *radar sites* are in *CC*?
2. What two types of *radar sites* are closest to the area designation "*M*"?
3. What are the two southernmost *radar sites* in *W*?

Locate

1. Which area(s) contain only one *headquarters*?
2. Which area(s) have two, and only two, *prisons*?
3. Which area(s) have four types of targets?

Count

1. How many *headquarters* are in *P, N, Y,* and *S*? .....
2. How many *barracks* are on the entire map? .....
3. How many *prisons* and *command posts* are in *DD*? .....

Compare

1. In area *AA*, are there more *prisons* than *command posts*? Yes No
2. Are there more *barracks* in *Q* than in *Y*? Yes No
3. In area *Z*, is a *headquarters* close to the area designator than a *barracks* is? Yes No

Verify

1. There are *two prisons* in *V*. T F
2. The *army personnel building* nearest the designation "*G*" is a headquarters. T F
3. Area *EE* contains *three* types of *army personnel buildings*. T F

ing could be expressed in terms of geometric means, rather than arithmetic means. It was believed that the geometric mean was superior to the arithmetic mean, because of the positive skewness of scores often encountered when time scores are involved.

Rate of responding for the Identify and Locate task was based on the number of items circled on the answer sheet. For the Count task, the total number of items counted was used. For the Compare task, the number of questions answered was used (i.e. three questions per trial). For the Verify task, the number of targets dealt with was used as a base.

Two major statistical analyses were conducted. An analysis of variance was performed to determine the effects of the experimental conditions on rate of responding. In addition, a factor analysis was conducted on the intercorrelations between the five tasks to identify some of the factors these tasks were tapping, and to obtain a better insight into the number and kind of tasks which might be used in future studies on visual displays for intelligence centers.

Effect of Experimental Conditions

A six-way mixed-model analysis of variance with subjects as the only random variable was conducted on the log of rate-of-response scores. The major results of this analysis are summarized in Table V. The table shows that all the main effects, except for Number of Targets, were significant at the 5 per cent level.

TABLE V

Summary of Analysis of Variance

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
Coding combinations (A)	14	0.1574	EA	2.19*
Number of levels (B)	1	9.1431	EB	139.06*
Clustering (C)	1	0.8969	EC	50.16*
Number of targets (D)	1	0.1418	ED	3.83
Subjects (E)	4	0.6601	Residual	66.25*
Task (F)	4	47.6228	EF	539.83*
BC	1	0.0997	BCE	2.99
BD	1	0.0187	BDE	—
BE	4	0.0658	Residual	6.60*
BF	4	1.4126	BFE	50.25*
BA	14	0.0546	BAE	—
CD	1	0.0519	CDE	12.43*
CE	4	0.0179	Residual	1.79
CF	4	0.5355	CFE	25.61*
CA	14	0.0054	CAE	—
DE	4	0.0370	Residual	3.71*
DF	4	0.2574	DFE	4.81*
DA	14	0.0065	DAE	—
EF	16	0.0882	Residual	8.85*
EA	56	0.7172	Residual	7.20*
FA	56	0.0776	FAE	7.01*
BCD	1	0.0668	BCDE	8.25*
BCE	4	0.0332	Residual	3.34*
BCF	4	0.0483	BCFE	1.79
BCA	14	0.0038	BCAE	—
BDE	4	0.0191	Residual	1.92
BDF	4	0.0184	BDFE	—
BDA	14	0.0066	BDAE	1.05
BEF	16	0.0281	Residual	2.82*
BEA	56	0.1590	Residual	15.96*
BFA	56	0.0086	BFAE	—
CDE	4	0.0042	Residual	—
CDF	4	0.1194	CDFE	16.20*
CDA	14	0.0044	CDAE	—
CEF	16	0.0209	Residual	2.10*
CEA	56	0.0065	Residual	—
CFA	56	0.0040	CFAE	—
DEF	16	0.0535	Residual	5.37*
DEA	56	0.0139	Residual	1.40*
DFA	56	0.0041	DFAE	—
EFA	224	0.0110	Residual	1.11
BCDE	4	0.0081	Residual	—
BCDF	44	0.0614	BCDFE	2.62
BCDA	14	0.0059	BCDAE	1.08
BCEF	16	0.0269	Residual	2.70*

Table V—continned

Source of variance	*d.f.*	*Mean square*	*Error term*	*F*
BCEA	56	0.0097	Residual	—
BCFA	56	0.0044	BCFBE	—
BDFA	56	0.0038	BDFAE	—
BDEA	56	0.0063	Residual	—
BEEF	16	0.0270	Residual	2.71*
BEFA	224	0.0127	Residual	1.27*
CDEF	16	0.0074	Residual	—
CDEA	56	0.0072	Residual	—
CDFA	56	0.0049	CDFAE	—
CEFA	224	0.0086	Residual	—
DEFA	224	0.0130	Residual	1.31*
BCDEF	16	0.0234	Residual	2.35*
BCDEA	56	0.0054	Residual	—
BDEFA	224	0.0091	Residual	—
CDEFA	224	0.0107	Residual	1.08
BCEFA	224	0.0079	Residual	—
BCDFA	56	0.0008	Residual	—
Residual	224	0.0100		

* $p > 0.05$.

TABLE VI

Duncan Multiple Range Tests Applied to Geometric Means of Response Rates for Each of Five Tasks

Rank Orders

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Identify:	20.65	20.59	19.93	19.80	19.53	17.21	16.81	16.60	16.55	16.53	16.33	16.27	16.09	15.99	14.95
Target	No.	No.	No.	No.	No.	Color	Color	Shape	Shape	Shape	Color	Shape	Color	Shape	Color
Background	All G.	Diff. G.	White	Patt.	Color	All G.	Patt.	White	Color	All G.	White	Diff. G.	Diff. G.	Patt.	Color
Locate:	9.38	9.36	9.19	905	8.95	7.67	7.48	7.30	7.16	7.07	6.87	6.82	6.77	6.70	6.54
Target	Color	Color	Color	Color	Color	No.	No.	No.	No.	No.	Shape	Shape	Shape	Shape	Shape
Background	Diff. G.	White	Patt.	All G.	Color	Patt.	Diff. G.	Color	White	All G.	All G.	Patt.	White	Diff. G.	Color
Count:	41.68	41.42	39.88	39.76	39.08	28.99	28.99	28.71	28.52	28.49	28.48	28.38	28.09	28.05	28.00
Target	Color	Color	Color	Color	Color	Color	No.	No.	Shape	Shape	Shape	No.	No.	No.	Shape
Background	Diff. G.	All G.	Patt.	White	Color	Diff. G.	Diff. G.	White	White	Color	All G.	All G.	Patt.	Color	Patt.
Compare:	7.54	7.25	7.13	7.13	7.10	7.09	7.02	7.01	6.98	6.96	6.96	6.87	6.66	6.58	6.54
Target	Color	No.	Color	Color	Shape	No.	Color	No.	No.	Color	Shape	No.	Shape	Shape	Shape
Background	Patt.	All G.	All G.	White	Diff. G.	Patt.	Diff. G.	Diff. G.	White	Color	All G.	Color	Color	White	Patt.
Verify:	20.37	20.20	20.13	20.12	20.08	20.06	19.65	19.64	19.32	19.21	19.06	18.89	18.86	18.72	18.24
Target	Color	Color	No.	No.	Shape	No.	Color	No.	No.	Shape	Color	Shape	Color	Shape	Shape
Background	All G.	White	All G.	Patt.	Diff. G.	Diff. G.	Diff. G.	White	Color	All G.	Patt.	White	Color	Patt.	Color

Notes:

(1) Numbers connected by line are not significantly different at $p < 0.05$.

(2) Abbreviations:
All G. = All gray.
Diff. G. = Different shades of gray.
Patt. = Patterned.
No. = Number.

Coding combinations. Primary interest in this analysis centers on the effects of the 15 coding combinations on operator performance, and on the consistency of these effects over different complexity and task conditions. Table V shows a significant main effect for coding combinations. The only significant interaction with coding, excluding interactions which include subjects, is the double interaction with task. Consequently, the coding combinations were not examined relative to the complexity conditions, but were examined only relative to task.

The significance of differences between coding combinations was tested for each task by the Duncan multiple range test.[5] The results are shown in Table VI. Each block of the table represents one of the five tasks. Within each block the coding combinations are listed horizontally in order of merit with the highest geometric mean response rate shown at the left, and the lowest shown at the right. Within each row, underline marks show those means which are not significantly different from one another at $p < 0.05$.

It is clear from Table VI that numbered targets are superior for the Identify task, and colored targets are superior for the Locate and Count tasks. For all task and target coding combinations, the background codings do not differ significantly; and for Compare and Verify, there are no significant differences among any of the 15 coding combinations.

Complexity conditions. There are no significant non-subject interactions between coding combinations and any of the complexity variables except the double interaction with task. When the density variables are considered, however, each density variable interacts significantly with the other two density variables, and each density variable interacts with task. Consequently, the effects of the density variables were studied in combination with themselves, and in combination with task. The data are summarized in Table VII.

In general, the effect of increasing the number of kinds of targets from four to eight is to decrease the rate of responding, as might be expected. This effect is especially pronounced in the case of the Count task.

Clustering of targets is related to an appreciable improvement in performance for only the Count task, especially with 120 targets, for both 4 and 8 levels of coding. A sizeable improvement also occurs in the Verify task under the 120-target, 4-levels condition. It appears

TABLE VII

Geometric Means of Responses per Minute for the Density Conditions, Broken Down by Task

	Identify					*Locate*			
	4–1/25	4–1/4	8–1/25	8–1/4		4–1/25	4–1/4	8–1/25	8–1/4
60	18.06	18.53	16.17	16.64	60	8.64	10.31	6.45	6.33
120	18.80	20.36	15.99	15.83	120	9.13	9.23	6.54	6.00
	Count					*Compare*			
	4–1/25	4–1/4	8–1/25	8–1/4		4–1/25	4–1/4	8–1/25	8–1/4
60	35.04	42.91	20.27	25.22	60	7.49	7.11	7.41	7.15
120	39.56	54.73	20.55	32.07	120	6.59	6.81	7.36	6.04

	Verify			
	4–1/25	4–1/4	8–1/25	8–1/4
60	20.52	19.93	16.43	17.77
120	18.85	25.28	18.11	20.02

that the clustered condition permits the operator to deal with target groups rather than with single targets. This would enable him to handle them at a faster rate, especially in counting, in which case he may count by 2's or 3's rather than by 1's.

It was most interesting to note that, under certain conditions, performance tends to improve as the total number of targets increases. This effect is found in approximately half the comparisons in Table VII. Considerable increases are found for Count with the exception of the 8–1/25 condition. An increase can be seen for the Verify task, except for 4 levels and clustering of 1/25.

Relation Between Tasks

It has been evident in all the experiments in the current research program that the particular task being performed by the operator is an important consideration in the evaluation of display characteristics. In the present experiment, the task variable not only had a significant main effect, but also entered into a number of significant interactions.

Because of the significant effect of the task variable, and because of its importance in the simulation of real operations, an additional analysis was made to determine some of the basic factors which account for variations in the five tasks. Thurstone's centroid method of factor analysis[9] was used to extract three factors. These three factors were rotated by graphic techniques to simple orthogonal structure, and one further rotation was made on the basis of enhancing psychological meaningfulness.

The data for the factor analysis are summarized in Table VIII. The matrix on the left-hand side of the table shows, above the diagonal, the original intercorrelation matrix for the five tasks (Identify, Locate, Count, Compare, and Verify). The tasks were correlated over the mean performance for each of the 120 display conditions for each task. Below the diagonal is the final residual matrix. The factor loadings and the communalities (h^2) for each task are shown in the four columns on the right-hand side of the table.

Factor I is identified as "Recognition". It appears to be a general factor, but with higher loadings on Identify and Verify, than on the other three tasks. The generality of this factor seems to be representative of the requirement of recognition, which is a necessary part of each task, but which is a larger part of Identify and Verify than of the remaining three tasks.

Factor II is identified as a "Search" factor. Its highest loadings are on Locate and Count. The Locate task requires the operator to scan the map and then check on a list the areas which contain certain types of targets. The Count task requires the operator to scan one area for two kinds of targets, four areas for one kind of target, or the whole map for one kind of target. It seems clear that these two tasks would have high loadings on a "Search" factor, and that the other tasks would have much

TABLE VIII

Intercorrelation Matrix, Residual Matrix, and Rotated Factor Loadings for Five Tasks

	*Intercorrelations and residuals**					*Task names*	*Rotated factors*			
	A	B	C	D	E		I	II	III	h^2
A		0.36	0.35	0.15	0.45	A (Identify)	0.64	0.09	0.22	0.47
B	−0.01		0.82	0.21	0.43	B (Locate)	0.26	0.68	0.62	0.91
C	−0.01	−0.01		0.06	0.67	C (Count)	0.40	0.89	0.12	0.97
D	−0.01	0.01	0.01		0.04	D (Compare)	0.15	−0.06	0.33	0.14
E	−0.01	−0.01	0.02	−0.01		E (Verify)	0.68	0.43	−0.08	0.65

* Correlations: upper right; residuals: lower left.

lower loadings. A moderately high loading for Verify, however, is not so apparently appropriate. This loading may be related to the fact that the Verify questions were all true–false questions with very few items (either targets or map areas) to be remembered or enumerated. All three questions referred to a single area each, with the area being designated within the question. Consequently, there can be very little else to this task beyond finding the appropriate area.

Factor III is named "Remembering", although it might have been called "Referral". Its highest loading is on the Locate task, which has already shown a high loading on the Search factor. It was not immediately apparent what the nature of this factor might be. A review of the operations associated with the performance of Locate, however, revealed a frequent shift of head position necessitated by the need to refer back to the question sheet periodically to check off areas from the thirty listed below the question. A moderate loading on the Compare task is interpreted as indicating the need for referring back to the questions, since every question dealt with two areas and one class of targets or with two classes of targets and one area.

The presence again, under altered conditions, of the Recognition and Search factors found in Experiment II increases confidence in the validity of these factors. Their basic meaningfulness is attested to by the finding that colored targets are superior for Locate and Count (high loadings on the Search factor) and numerical targets are superior for Identify (high loadings on the Recognition factor). There are no differences in performance, so far as colored targets versus numeral targets is concerned, for Compare (no appreciable correlation with either Search or Recognition) and for Verify (high loadings on both factors).

The third factor, Remembering, seems to be related more to the structure of the operator's work, and to the quantity of data to be dealt with at one time, than to the design of the display unit. It suggests further research on time-saving features for use when many items must be recorded, or when different kinds of data must be studied at the same time.

Preferences

Preference ratings for the 15 coding combinations were obtained from the five subjects at the conclusion of the experiment. These ratings were converted to rank-order data, and Friedman's chi square[9] was used to test the significance of differences among ranks for the total table. This test showed that the 15 combinations as a whole differed. The same test was then applied separately to background codes and to target codes. Background codes did not show a significant difference, but target codes did.

A summary of the preference rankings is presented in Table IX. A cursory examination of the table reveals the subjects' pronounced preference for colored targets. It will be recalled that the colored targets were associated with superior performance on the Locate and

TABLE IX

Preferences by Ranks Averaged Over Five Subjects for Fifteen Code Combinations

Targets	*Backgrounds*					*Average rank over backgrounds*
	Color	*Mixed grays*	*Solid gray*	*White*	*Patterned*	
Color	6.8	3.5	4.1	3.5	3.9	4.36
Number	8.4	9.4	8.8	9.8	13.5	9.98
Enclosed shape	6.8	9.6	10.2	9.6	12.1	9.66
Average rank over targets	7.33	7.50	7.70	7.63	9.83	

Count tasks, both of which were highly correlated with the Search factor. These two tasks were also the two longest in terms of absolute time for completion. Consequently, it is no wonder the subjects preferred the attention-getting colored targets when so much of their time was spent on the Search type of task. Had the total task been more heavily loaded with Identify questions (and the Recognition factor) it is possible that the preferences might have been for numerals rather than for colored targets.

DISCUSSION AND CONCLUSIONS

In the interpretation of the results of this study, the particular circumstances of the experiment must be kept in mind. A limited scope as to the kinds of coding combinations studied, the types and degrees of complexity, the structure of the operator's task, and the range of coding techniques naturally restricts the generalizability of results.

Operator Tasks

The question of what kind of task to give the operator in the experimental setting is of great importance. The task selection of this study was based on the experimenters' ideas about what tasks would be involved in working with a cartographic display. A study of actual operational activities would enable future experimenters to incorporate tasks that would better simulate reality. This is a worth-while goal in view of the significant interactions between tasks and coding techniques.

The importance of selecting appropriate operator tasks is further pointed up by the unusual finding that for some tasks an increase in the total number of targets shown on the display results in a faster rate of performance. This result did not occur for all tasks, however. It is a result that never would have been found in many visual display studies, since such studies are based on tasks such as identifying a single flash target as quickly as possible, or finding a previously known and unique target which is embedded in a mass of other targets.

The factor analysis of tasks must be viewed with certain limitations. First, it would be desirable to intercorrelate a much larger number of tasks, representing a wider sampling of human-operator functions on visual displays. Second, it would be better to correlate tasks over a wider range of visual display conditions. Both of these improvements would probably result in a larger number of factors, with better definition and interpretation of each factor. It would then be possible to develop tasks which are reasonably "pure" factorially. In turn, such tasks would result in:

(1) A more meaningful and generalizable interpretation of the effects of display variables on performance.
(2) A more meaningful interpretation of task-display parameter interactions.

Subject Differences

The interactions between subjects and display parameters indicate individual differences in responses to coding techniques and to the complexity of visual displays. While the nature and extent of such differences were not studied in detail in this experiment, the significant interactions warrant further study. For example, there is a definite possibility that a coding scheme which is best as judged by averages taken over a number of operators, may not be best for certain individual operators. If this were found to be the case, it would be possible for the wrong operator to be on duty at a crucial time.

Color Coding

Because of the practical problems of generating and transmitting color-coded displays, and because of a general belief that color coding is desirable, there has been a particular interest in the relative efficiency of performance with color codes. For the Identify task, number coding was superior to color coding, which is in agreement with Hitt's findings. For both Locate and Count, however, color coding was superior to number; but in Hitt's study, no

significant differences were found between color and number coding for these two tasks. Perhaps the discrepancy can be explained in part by the greater display complexity involved in the present experiment, which would tend to enhance the attention-getting qualities of color. Also, there were certain differences between the two studies with respect to the type of Count and Locate questions used.

With regard to background, colored backgrounds did not differ significantly from non-colored backgrounds. In addition, the four black-and-white backgrounds did not differ significantly among themselves. Thus, it would appear that the choice of background would depend upon such factors as ease of construction and economy.

In favor of color coding is the fact that numbers might be held in reserve to be used for numerosity indicators, possibly in conjunction with some other code. Arguments against the use of color include: (1) the possibility of color-deficient vision among operators; and (2) a number of unanswered questions; for instance, there may be other untried codes that would do as well.

The following conclusions apply only to the conditions used in this study:

(1) There are no significant differences between the selected background codes.
(2) The choice of target codes depends upon the task to be performed.
(3) For targets, color coding is superior for Locate and Count tasks; numeral coding is superior for the Identify task.
(4) Subjects preferred color targets to numeral and shape targets.
(5) Coding combinations and complexity conditions do not interact.
(6) Intercorrelations between the five tasks included in the study may be explained in terms of three factors: Recognition, Search, and Remembering.

REFERENCES

1. Baker, C. A., and Grether, W. F., *Visual Presentation of Information*, WADC Technical Report 54–160, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

2. Bridgman, C. S., and Wade, E. A., "Optimum Letter Size for a Given Display Area". *J. Appl. Psychol.*, **40**, 378–380, 1956.

3. Christner, Charlotte A., Schutz, H. G., and Ray, H. W., *Some Factors Affecting Visual Search Time for Symbols on a Large Visual Display*. Paper presented at the Annual Convention, American Psychological Association, Cincinnati, Ohio, September, 1959.

4. Conover, D. W., and Kraft, C. L., *The Use of Color in Coding Displays*, WADC Technical Report 55–471, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, August, 1954.

5. Duncan, D. B., "Multiple Range and Multiple F Tests". Reprint Series No. 49; reprinted from *Biometrica*, **11** (1), 1955.

6. Fitts, P. H., *Stimulus Characteristics Determining Speed of Classifying Visual Patterns*. Minutes and Procedures of 35th Meeting of Armed Forces–NRC Vision Committee (1954), Ann Arbor, Michigan; Armed Forces–NRC Vision Committee Secretariat, University of Michigan, pp. 52–60, 1959.

7. Hitt, W. D., "Experiment IV: An Evaluation of Five Different Abstract Coding Methods" (present edition of this journal).

8. Lindquist, E. F., *Design and Analysis of Experiments in Psychology and Education*, Houghton Mifflin Company, Boston, 1953.

9. Wilcoxon, F., *Some Rapid Approximate Statistical Procedures*, American Cyanamid Company, New York, 1949.

The Authors

DARWIN P. HUNT (*The Effect of the Precision of Informational Feedback on Human Tracking Performance*) is a research psychologist at the Engineering Psychology Branch, Behavioral Sciences Laboratory, Wright Air Development Division. He has served in this capacity since 1951 and has published research in the areas of control coding, tactual guidance, audio warning systems, learning theory and human tracking dynamics.

Dr. Hunt received his B.A. degree (mathematics and psychology) from Miami University (Ohio) in 1950. He was awarded his Ph.D. degree (experimental psychology and statistics) by Ohio State University in 1960. Dr. Hunt is currently responsible for a research program directed toward determining the influence of transformations of displayed error information on the ability of humans to perform various tasks such as controller functions.

W. D. HITT, H. G. SCHUTZ, C. A. CHRISTNER, H. W. RAY and J. L. COFFEY (*Development of Design Criteria for Intelligence Display Formats*)

JOHN L. COFFEY (*A Comparison of Vertical and Horizontal Arrangements of Alphanumeric Material [Experiment I]*) is a member of the human factors research group of Battelle's Systems Engineering Division. His major interests are in experimental and engineering psychology. Mr. Coffey received his B.S. and M.S. degrees in psychology from the Ohio State University, where he is presently working toward his Ph.D. degree in psychology. When Mr. Coffey joined Battelle in 1955, he worked as a senior technician in non-ferrous physical metallurgy. Since he joined the human factors group in September, 1959, he has been engaged in research on large-scale visual displays and research on aural reading devices. His work on visual displays involved the design, conduct, and analysis of experiments on alphanumeric presentation of intelligence information. In the area of aural reading devices, he has actively participated in the development of improved hardware, better tracking techniques, and improved training procedures.

HOWARD G. SCHUTZ (*An Evaluation of Formats for Graphic Trend Displays [Experiment II] and an Evaluation of Methods for Presentation of Graphic Multiple Trends [Experiment III]*) is a senior psychologist in the biosciences research group of Battelle Memorial Institute. His major interests are in physiological and experimental psychology. He did his undergraduate work at the University of Illinois and received his M.S. and Ph.D. degrees in physiological and experimental psychology from the Illinois Institute of Technology in 1952 and 1955, respectively. Before joining Battelle in 1957, Dr. Schutz served for five and a half years as a research psychologist for the Quartermaster Food and Container Institute. His research activities were in the areas of taste, odor, appetite, and food preferences. Since joining the staff of Battelle, Dr. Schutz has worked on projects of design of visual displays, console layout, basic mechanisms of olfaction, and sensory evaluation of products. Dr. Schutz is a member of many technical and professional societies, including the American Psychological Association and the Human Factors Society. He is a member of Sigma Xi. He has published more than twenty technical reports and articles in the general area of applied experimental psychology.

WILLIAM D. HITT (*An Evaluation of Five Different Abstract Coding Methods [Experiment IV]*). See page 171, *Human Factors*, August, 1960, Vol. 2, No. 3.

CHARLOTTE A. CHRISTNER and HORACE W. RAY (*An Evaluation of the Effect of Selected Combinations of Target and Background Coding on Map-reading Performance [Experiment V]*)

CHARLOTTE A. CHRISTNER, Project Leader, Systems Engineering Division, is an experimental psychologist with major interests in visual displays and information handling in large-scale systems. She is a graduate of Miami University and the Ohio State University. Dr. Christner joined Battelle in 1956. Prior to joining the Battelle staff, Dr. Christner had experience in a variety of psychology problems, including motor skills, human engineering, and statistical analysis. One of her first interests at Battelle was in systems research on electronic countermeasures. While she has assisted on a variety of projects, her work has been concerned primarily with human-factors problems of large semi-automatic systems. In particular, this has focused on studies on visual displays and on information handling. Dr. Christner is a member of the Human Factors Society, of the Central Ohio Psychological Association (a member of the board of directors), of The American Psychological Association (Division 8—Personality and Social Psychology, and Division 21—The Society of Engineering Psychologists), and of the Highway Research Board.

HORACE W. RAY. See page 172, *Human Factors*, August, 1960, Vol. 2, No. 3.

EDITOR'S NOTE: The following three biographies were received too late for inclusion in Vol. 3, No. 1, March 1961 issue.

ROBERT H. JOHNSON, DONALD A. GORDON, BRUCE O. BERGUM, and WAYNE E. PATTERSON (*COED—A Device for the Experimental Study of Man-Machine Systems*)

ROBERT H. JOHNSON is Assistant Department Head, Data Processing and Systems Department, Bendix Systems Division, Ann Arbor, Michigan. In this position, he has made substantial contributions to the overall systems design of the EAGLE, and other contemporary missile systems. Prior to working at Bendix, he held positions at Pratt and Whitney, and at the RADC missile support laboratory. He received his B.S. at the University of Illinois, and holds masters degrees in aeronautical and instrumental engineering (both 1954) from the University of Michigan.

DONALD A. GORDON is Head of the Experimental Simulation Group, Data Processing and Systems Department, Bendix Systems Division, Ann Arbor, Michigan. Prior to his present position, he was assistant professor of psychology at the University of Illinois, and held positions in the Personnel Research Branch, TAGO, and in the Engineering Research Institute, University of Michigan. He received his Ph.D. in psychology in 1947, from Columbia University.

BRUCE O. BERGUM is a research scientist at the Human Research Unit (HumPRO), Fort Bliss, Texas. He was formerly Head, Human Factors Group, Data Processing and Display Department, Bendix Systems Division, and has held teaching positions with Northwestern University and the University of Illinois. Dr. Bergum has worked in training and maintenance research for the Air Force Personnel and Training Research Center and is currently engaged in research on missile system operator problems. He received his M.S. in experimental psychology in 1953 and his Ph.D. in 1958 from Northwestern University.